FORM-PTO-1390 (Rev. 12-29-99) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER: 012627-019

U.S. APPLICATION NO. (If known, see 37 C.F.R. 1.5): 09/720215

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/DE99/01867	25 June 1999	26 June 1998

TITLE OF INVENTION: MODULARLY CONSTRUCTED RNA MOLECULES HAVING TWO SEQUENCE REGION TYPES

APPLICANT(S) FOR DO/EO/US: Annemarie POUSTKA; Johannes COY

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.

2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.

3. ☒ This is an express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and the PCT Articles 22 and 39(1).

4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.

5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))

   a. ☒ is transmitted herewith (required only if not transmitted by the International Bureau).

   b. ☒ has been transmitted by the International Bureau.

   c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US)

6. ☒ A translation of the International Application into English (35 U.S.C. 371(c)(2)).

7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))

   a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).

   b. ☐ have been transmitted by the International Bureau.

   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.

   d. ☐ have not been made and will not be made.

8. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).

9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).

10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11. to 16. below concern other document(s) or information included:

11. ☒ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.

12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.

13. ☒ A FIRST preliminary amendment.

    ☐ A SECOND or SUBSEQUENT preliminary amendment.

14. ☐ A substitute specification.

15. ☐ A change of power of attorney and/or address letter.

16. ☐ Other items or information:

U.S. APPLICATION NO. (if known) see 37 C.F.R. 1.50)	INTERNATIONAL APPLICATION NO.	ATTORNEY'S DOCKET NUMBER
09/720215	PCT/DE99/01867	012627-019

17. ☒ The following fees are submitted:	CALCULATIONS	PTO USE ONLY
Basic National Fee (37 CFR 1.492(a)(1)-(5)):		
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO $1,000.00 (960)		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO $860.00 (970)		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO $710.00 (958)		
International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4) $690.00 (956)		
International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4) $100.00 (962)		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 860.00	
Surcharge of **$130.00 (154)** for furnishing the oath or declaration later than 20 ☐ 30 ☐ months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

Claims	Number Filed	Number Extra	Rate		
Total Claims	26 -20 =	6	X$18.00 (966)	$ 108.00	
Independent Claims	3 -3 =	0	X$80.00 (964)	$ --	
Multiple dependent claim(s) (if applicable)			+$270.00 (968)	$	
TOTAL OF ABOVE CALCULATIONS =				$ 968.00	
Reduction for 1/2 for filing by small entity, if applicable.				$ 484.00	-
SUBTOTAL =				$ 484.00	
Processing fee of **$130.00 (156)** for furnishing the English translation later than 20 ☐ 30 ☐ months from the earliest claimed priority date (37 CFR 1.492(f)).			+	$	
TOTAL NATIONAL FEE =				$ 484.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00 (581)** per property +				$	
TOTAL FEES ENCLOSED =				$ 484.00	
				Amount to be: refunded	$
				charged	$

a. ☒ A check in the amount of $ 484.00 to cover the above fees is enclosed.

b. ☐ Please charge my Deposit Account No. 02-4800 in the amount of $_____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 02-4800. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Teresa Stanek Rea
BURNS, DOANE, SWECKER & MATHIS, L.L.P.
P.O. Box 1404
Alexandria, Virginia 22313-1404
(703) 836-6620

SIGNATURE

Teresa Stanek Rea
NAME

30,427
REGISTRATION NUMBER

Patent
Attorney's Docket No. 012627-019

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of	
Annemarie POUSTKA et al.	
Application No.: Unassigned (Corresponds to PCT/DE99/01867)	Group Art Unit: Unassigned
International Filing Date: 25 June 1999	Examiner: Unassigned
For: MODULARLY CONSTRUCTED RNA MOLECULES HAVING TWO SEQUENCE REGION TYPES	

PRELIMINARY AMENDMENT

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Prior to examination, please amend the above-captioned application as follows:

IN THE CLAIMS:

Kindly amend the claims as follows:

Claim 3, line 1, delete "or 2".

Claim 4, line 1, change "any one of claims 1 to 3" to --claim 1--.

Claim 5, line 1, change "any one of claims 1 to 4" to --claim 1--.

Claim 6, line 1, change "any one of claims 1 to 5" to --claim 1--.

Claim 7, line 2, change "any one of claims 1 to 6" to --claim 1--.

Claim 9, line 2, delete "or the gene according to claim 8".

Claim 14, lines 1-2, change "any one of claims 9 to 13" to --claim 9--.

Claim 16, lines 2-3, change "any one of claims 1 to 6" to --claim 1--.

Claim 18, line 2, change "any one of claims 1 to 6" to --claim 1--.

Claim 19, line 2, change "any one of claims 1 to 6" to --claim 1--.

20. (Amended) [Use of the RNA molecule according to any one of claims 1 to 6, of the vector according to any one of claims 9 to 13, of the antibody or fragment thereof according to claim 16 or 17, of the antisense RNA according to claim 18 or of the ribozyme according to claim 19 for the production of a] A pharmaceutical preparation for preventing or treating diseases which are connected with a disturbed control of gene expression comprising using the RNA molecule according to claim 1.

21. (Amended) [Use of the RNA molecule according to any one of claims 1 to 6, of the DNA sequence according to claim 7 or a fragment thereof, of the antibody or fragment thereof according to claim 16 or 17, or of the antisense RNA according to claim

18 or a fragment thereof] A method for the diagnosis of diseases which are connected with a disturbed control of gene expression comprising using the RNA molecule according to claim 1.

Claim 22, line 1, change "Use" to --The method-- and delete "20 or".

Claim 23, line 1, change "whose" to --comprising a-- and after "gene" insert --which--.

Claim 25, line 1, delete "or 24".

26. (Amended) A process for the production of a non-human mammal according to [any one of claims 23 to 25] claim 23, [characterized by] comprising the following steps:

(a) [preparation of] preparing a DNA fragment, [in particular a vector,] containing a modified NINTROX gene, the NINTROX gene having been modified by deletion of a homologous sequence and/or insertion of a heterologous sequence[, in particular a selectable marker];

(b) [preparation of] preparing embryonal stem cells from a non-human mammal [(preferably mouse)];

(c) [transformation of] transforming the embryonal stem cells from step (b) with the DNA fragment from step (a), the NINTROX gene in the embryonal stem

cells being modified by homologous recombination with the DNA fragment from (a),

(d) culturing the cells from step (c),

(e) [selection of] selecting the cultured cells from step (d) for the absence of the homologous sequence and/or the presence of the heterologous sequence, [in particular the selectable marker,]

(f) [production of] producing chimeric non-human mammals from the cells from step (e) by injection of these cells in mammalian blastocysts [(preferably mouse blastocysts)], [transfer of] transferring the blastocysts into false-pregnant female mammals [(preferably mouse)] and [analysis of] analyzing the resulting offspring for a change of the NINTROX gene.

REMARKS

Entry of the foregoing amendments are respectfully requested.

Should the Examiner have any questions concerning the subject application, a telephone call to the undersigned would be appreciated.

Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

By: Teresa Stanek Rea
Teresa Stanek Rea
Registration No. 30,427

P.O. Box 1404
Alexandria, Virginia 22313-1404
(703) 836-6620

Date: December 22, 2000

Modularly Constructed RNA Molecules Having two Sequence Region Types

The present invention relates to RNA molecules which are characterized by two sequence region types, namely a first sequence region type which contributes to maintaining the three-dimensional structure of the RNA molecule, and a second sequence region type which is responsible for the specific binding of a ligand. These RNA molecules are preferably useful for the direct control of gene expression. The present invention also provides the DNA sequence derived for the RNA molecules according to the invention and vectors which contain them. In addition, the invention relates to drugs or medicaments and diagnostic compositions which contain the above RNA molecules or vectors, to an antibody specifically recognizing these RNA molecules or to antisense RNA specifically binding to these RNA molecules or ribozymes cleaving these RNA molecules. Furthermore, the invention relates to non-human transgenic mammals and cells obtained therefrom.

Gene expression in eukaryotes is usually regulated via proteins which usually bind specifically to certain regulatory sequences upstream of the gene to be expressed and show a characteristic effect (RNA polymerases, transcription factors, receptors adapted to be activated by hormones, etc.). Only few examples of controlling the gene expression directly via RNA molecules have been known thus far. They include the RNA "XIST" responsible for the inactivation of the entire X chromosome ("X chromosome inactivation specific transcript"), an RNA referred to as

IPW ("imprinted in Prader-Willi syndrome") and RNA H19 which represents a tumor suppressor and is involved in the control of certain development processes. The artificial control of the gene expression has meanwhile been effected by the use of antisense RNAs binding specifically to mRNAs or by the use of catalytically active RNA molecules, what is called ribozymes, which do not only bind specifically to the target RNA but also cleave it thus inactivating it. However, the application possibilities for these antisense RNAs or ribozymes are limited, above all as regards the ligand to be bound and inactivated. This ligand may basically only be an RNA.

Thus, there is a need for providing compounds which can universally detect, and/or inactivate, the most differing target molecules, e.g. DNA, RNA, proteins or low-molecular substances, and are suitable e.g. for controlling gene expression and thus, of course, also for preventing and treating diseases which are accompanied by a disturbed gene expression.

Hence the technical problem of the invention is substantially to provide those compounds which are useful *inter alia* for the prevention or therapy (and also diagnosis) of such diseases.

The solution to this technical problem was achieved by providing the embodiments characterized in the claims.

The inventors could identify an RNA molecule which comprises the above described desired properties. This RNA molecule is encoded by the gene "NINTROX" (No INTROns X-chromosome) which has no introns, is localized on the X-chromosome and codes for no protein. This RNA molecule is part of certain

(relatively long) transcripts of the MeCP2 gene. The MeCp2 gene (methyl-CpG binding protein 2) in Xq28 has a transcript of about 1.8 kb which codes for the MeCP2 protein. The above described RNA is part of relatively long MeCP2 transcripts which also code for the MeCP2 protein but have a different 3'-non-translated region. This 3'-non-translated region is decisive for the MeCP2 gene and its function. The below expression "NINTROX" is synonymous with the above relatively long transcripts of the MeCP2 gene.

The genomic sequence of the human NINTROX gene is shown in figure 1, and the genomic sequence of the murine NINTROX gene is illustrated in figure 2. In figure 3, a sequence comparison was carried out between human and murine sequences. It is obvious therefrom that there are some highly sequence-conserved regions which according to an energy analysis carried out by means of a computer distinguish themselves by a high degree of energy (cf. figure 4).

While the mechanism of action of the above discussed genes effective on the RNA level was fully unclear, the principle of action of such a gene which is described in more detail below could, for the first time, be determined by the analysis of the NINTROX gene. The NINTROX gene contributes essentially to the maintenance of the functions of the CNS, in particular the hippocampus. Defects in this gene result in limited CNS functions which reach as far as mental retardations. Furthermore, the NINTROX gene has an important function in the control of cell proliferation. In this connection, changes in this gene can lead to errors in the control of cell growth, e.g. to cancer. Changes in this gene may result in an increased or reduced DNA methylation. An increased DNA methylation can *inter alia* restrict or prevent

the activity of growth-controlling genes (tumor suppressor genes) and thus result in a generally increased cancer rate. Reduced DNA methylation can lead *inter alia* to an overexpression of genes and thus to a disturbed development of the cell or the whole organism. Further investigations led to the result that the expression pattern of the NINTROX gene is effected in tissue-specific and development-specific manner. The Northern analyses showed an expression in all investigated fetal and adult tissues. No sequence homologies with already known sequences could be detected.

The strategy which led to the identification of this nucleic acid molecule is described below. Based on the systematic analysis of the q28 region of the human X chromosome various expressed sequences could be detected and isolated. By means of these expressed sequences some formerly unknown genes could be identified and characterized according to standard methods, *inter alia* the NINTROX gene on which the present invention is based.

It is of interest that the NINTROX-RNA molecules according to the invention have a modular structure, i.e. they are characterized by the presence of two different sequence region types. While one sequence region permits to maintain the three-dimensional structure and, as follows from a comparison of the sequences from various species (human, hamster, kangaroo, macaque or macaca, orangutan chimpanzee and rat; cf. figure 5), is conserved only in a qualified sense, the second sequence region which is responsible for the specific binding to the target molecule is sequence-conserved. Because of this modular construction of the NINTROX-RNA it is possible to modify it such that its effect is not only limited to the above described control of the gene expression but can be used for a number of

possibilities. In addition to the control of the gene expression it is also possible to modify the structure (e.g. chromatin structure, nuclear scaffold) of chromosomal regions by means of such modular RNA molecules. This offers the formerly unknown possibility of being able to influence the expression of relatively large genomic regions in well-calculated fashion. Thus, certain sequence regions of both modules of the NINTROX gene can be replaced by other sequences or even artificial sequences, so that (a) the interaction of this RNA with other binding partners (RNA, DNA, other macromolecules and low-molecular compounds) or their biochemical reaction (e.g. increase or decrease of the conversion rate) are changed in well-calculated fashion, and therefore the RNA molecule can be adapted in well-calculated fashion to novel tasks, and/or (b) the three-dimensional structure of the NINTROX-RNA can be adapted in well-calculated fashion to special demands. As a result, a partially or fully new function of the NINTROX-RNA molecule according to the invention can be obtained.

Thus, an embodiment of the present invention relates to an RNA molecule which may bind to a ligand and comprises the following sequence regions: (a) a sequence region maintaining the three-dimensional structure of the RNA molecule, and (b) a sequence region for the specific binding of the ligand.

The expression "a sequence region maintaining the three-dimensional structure of the RNA molecule" used herein has the following meaning. Three-dimensional RNA structures are rendered possible by base pairing of various bases within the RNA molecule. In this case, structures such as "stems" or "loops" are formed. Many of these structures yield in this way the overall structure of the RNA molecule. A

sequence change within the RNA molecule may remain without consequences for the spatial structure if the sequence change does not change the base pairings or if the sequence change is compensated by a second sequence change. For example, if the base pairing A-T is destroyed in that the A mutates into G, this mutation can be compensated by another mutation of T into C. Although this changes the sequence, the spatial structure remains the same. As a result, the same RNA structure can be formed by an extremely large number of differing RNA sequences. References to certain RNA structures follow from an analysis of the energy included therein. This analysis can be carried out by means of commercially available computer programs (e.g. "FOLD"; Michael Zuker and P. Stiegler: Optimal Computer Folding of Large RNA Sequences Using Thermodynamics and Auxiliary Information, Nucleic Acids Research (81), 9(1), page 133). The lower the energy content of a certain sequence, the more stable the three-dimensional RNA structures. The analysis of the NINTROX gene showed a conserved distribution of these low-energy structures (figure 4). The base sequence of these RNA regions differs widely with various species, but the energy content is very conserved. In figure 3, these are the sequence regions which are not characterized by a black bar at the margin. This means that the sequence region maintaining the three-dimensional structure of the RNA molecule is not sequence-conserved but energy-conserved. For example, modifications of this sequence region do not orient themselves by the base sequence but by the conservation of the detected energy content.

The expression "a sequence region for the specific binding of the ligand" used herein relates to a sequence region which is such that it can bind specifically the desired ligand. These sequence regions are highly sequence-

conserved. In figure 3, these regions are marked by a black bar at the margin and have a high energy content (cf. figure 4). This tallies with the observation that these sequence regions are not "packed" but oriented outwardly and are responsible for the binding of the ligand, enzymatic reactions or the binding to other RNA or DNA sequences. If the ligand to be bound is an RNA molecule or a DNA molecule, this sequence region will be complementary to a corresponding, sufficiently long segment of the RNA molecule or DNA molecule. If the ligand to be bound is a protein, the sequence region (b) may be partially or fully exchanged, or supplemented, by a DNA sequence which as is known binds specifically the desired protein.

The two above-described sequence types occur several times within the NINTROX-RNA. The exchange or the change of individual ones of such modules enables the well-calculated change of the NINTROX-RNA. In a modification of the module maintaining the three-dimensional structure attention has to be paid to the energy content, so that it maintains a minimum value. The modification of the other sequence region is only subject to minor restrictions even though it is deemed to be sequence-conserved. This region may be omitted fully or partially or may contain insertions. For example, it is also possible to insert sequences into the NINTROX-RNA molecule which have known biochemical properties or bind certain DNA molecules, RNA molecules or proteins. In addition, random sequences of differing length may be introduced into various sites of the NINTROX gene and thereafter selection for specific properties such as biochemical reaction, specific binding, etc., may be carried out.

In a preferred embodiment of the RNA molecule according to the invention the sequence region (a) comprises the sequence regions not marked at the margin in figure 3 or sequences related thereto which also permit the maintenance of the three-dimensional structure of the RNA molecule and differ from sequence region (a) in figure 3. These differences relate to the addition, deletion and/or insertion of bases, at least 80 %, preferably 85 %, and more preferably at least 90 %, of the energy content determined for the sequence of figure (3) being maintained. The original three-dimensional structure is preferably maintained when these changes are introduced.

In a particularly preferred embodiment, the sequence region (b) of the RNA molecule according to the invention comprises the sequences which are illustrated in figure 3 and marked with black bars at the margin.

In another preferred embodiment of the RNA molecule according to the invention, the ligand to be bound is a DNA molecule or a protein or enzyme, e.g. DNA polymerase I. The RNA molecule according to the invention preferably contains a poly(A) sequence at the 3' end, which may contribute to the stability in a desired host cell.

In another preferred embodiment, the RNA molecule according to the invention is used to control the gene expression. For this purpose, the sequence region (b) is modified such that it binds a protein responsible for gene expression or binds to a certain DNA region of the target gene so as to impede or prevent e.g. the attachment of proteins which exert an influence inhibiting or supporting gene expression or also binds directly to the mRNA of the target gene so as to impede or prevent the translation, for example. The person

skilled in the art can readily modify the RNA molecule according to the invention by corresponding modification of sequence region (b) and possibly also of sequence region (a) such that it binds the desired ligand and therefore controls the gene expression to the desired extent.

The present invention also relates to a DNA sequence coding for the RNA molecule according to the invention and to a gene comprising the following features: It contains a promoter which permits the transcription in a desired host cell and a DNA sequence functionally linked therewith and encoding the RNA molecule according to the invention. The gene preferably contains additionally a termination signal and a polyadenylation site.

In a preferred embodiment, the gene according to the invention comprises the sequence shown in figure 1 or 2.

The DNA sequences or genes, coding for the RNA molecule according to the invention, may also be inserted in a vector. Thus, the present invention also comprises vectors containing these DNA sequences or genes. The term "vector" relates to a plasmid (e.g. pUC18, pBR322, pBlueScript), to a virus or another suitable vehicle. In a preferred embodiment, the sequence coding for the DNA molecule according to the invention is functionally linked in the vector with regulatory elements which permit its expression in prokaryotic or eukaryotic host cells. In addition to the regulatory elements, e.g. a promoter, such vectors typically contain a replication origin and specific genes which permit the phenotypic selection of a transformed host cell. The regulatory elements for the expression in prokaryotes, e.g. *E. coli*, comprise the lac, trp promoter or T7 promoter, and those for the expression in eukaryotes comprise the AOX1 or

GAL1 promoter in yeast and those for the expression in animal cells comprise the CMV, SV40, RVS-40 promoter, CMV or SV40 enhancer. Further examples of suitable promoters are the metallothionein I and the polyhedrin promoters. Suitable vectors are e.g. expression vectors, based on T7, for the expression in bacteria (Rosenberg et al., Gene 56 (1987), 125), pMSXND for the expression in mammalian cells (Lee and Nathans, J. Biol. Chem. 263 (1988), 3521) and vectors derived from baculovirus for the expression in insect cells.

In a preferred embodiment, the vector containing the sequences coding for the RNA molecules according to the invention is a viral vector, e.g. a vaccinia virus or adenovirus, which is of use for a gene therapy. RNA viruses, above all retroviruses, are particularly preferred. Examples of suitable retroviruses are MoMuLV, HaMuSV, MuMTV, RSV or GaLV. For the purpose of gene therapy the RNA molecules according to the invention can be transported to the target cells in the form of colloidal dispersions as well. They comprise e.g. liposomes of lipoplexes (Mannino et al., Biotechniques 6 (1988), 682).

General methods known in the art can be used for constructing expression vectors which contain the sequences coding for the RNA molecules according to the invention and suitable control sequences. These methods comprise e.g. *in vitro* recombination techniques, synthetic methods and *in vivo* recombination methods, as described in Sambrook et al., for example.

The present invention also relates to host cells containing the above described vectors. These host cells comprise bacteria, yeast, insect and animal cells, preferably mammalian cells. Preferred mammalian cells are CHO, VERO,

BHK, HeLa, COS, MDCK, 293 and WI38 cells. Methods of transforming these host cells, of phenotypically selecting transformants and expressing the nucleic acid molecules according to the invention using the above described vectors are known in the art.

The present invention also relates to antibodies which detect specifically the RNA molecule according to the invention. The antibodies may be monoclonal, polyclonal or synthetic antibodies or fragments thereof, e.g. Fab, Fv or scFv fragments. In this case, a monoclonal antibody is preferably concerned. The antibodies according to the invention may be produced according to standard methods, the RNA molecule according to the invention or a fragment thereof serving as an immunogen. Monoclonal antibodies may be produced e.g. by the method described by Köhler and Milstein (Nature 256 (1975), 495) and Galfré (Meth. Enzymol. 73 (1981), 3), mouse myeloma cells being fused with immunized mammalian spleen cells. These antibodies may be used e.g. to inhibit the activity of the RNA molecules according to the invention, e.g. to influence the gene expression. The antibodies may also be used in diagnostic assays, for example, so as to prove whether dysregulation of the gene expression is accompanied e.g. by a loss or lack of responsible NINTROX-RNA. The antibodies may be present in immunoassays in liquid phase or be bound to a solid carrier. In this connection, the antibodies may be labeled in various ways. Suitable markers and labeling methods are known in the art. Examples of immunoassays are ELISA and RIA.

The invention also relates to antisense RNAs which bind specifically to an RNA molecule according to the invention and may be used *in vitro* or *in vivo* to reduce the expression of genes controlled directly by RNA, e.g. NINTROX-RNA. The

administration of the antisense RNA according to the invention to a target cell results in a reduced gene expression and is particularly useful for treating diseases which are characterized by an excessively great gene expression of the directly RNA-controlled gene (e.g. cancer diseases). In this connection, the antisense RNAs can be administered directly or as a DNA encoding the same, preferably inserted in a suitable vector. The suitable vectors comprise all of the vectors described above already in connection with the RNA molecules according to the invention.

The antisense RNAs according to the invention comprise an antisense sequence having at least 7 to 10 or more nucleotides which hybridize specifically with a sequence of the RNA molecule according to the invention, e.g. NINTROX-RNA. The antisense RNA according to the invention preferably has a length of about 10 to about 50 nucleotides or of about 14 to about 35 nucleotides. In further embodiments, the antisense RNAs according to the invention are RNAs shorter than about 100 nucleotides or shorter than about 200 nucleotides. In general, the antisense RNAs should be long enough to form a stable double helix but short enough (depending on the kind of supply) to be administered *in vivo*, if desired. In general, the antisense sequence is substantially complementary to the target sequence to ensure specific hybridization. In certain embodiments the antisense sequence is directly complementary to the target sequence. However, the antisense RNAs may also contain nucleotide substitutions, additions, deletions, transitions, transpositions or modifications as long as the specific bond to the relevant target sequence is maintained as a functional property of the antisense RNA. The antisense RNAs may also contain further sequences in addition to the

antisense sequences. The antisense RNAs (and the RNA molecules according to the invention) can be produced using any method suitable for the production of nucleic acids, e.g. by chemical synthesis *de novo* or by cloning. An antisense RNA may also be produced e.g. by inserting in a vector (e.g. a plasmid) a sequence of the target RNA or a fragment thereof in reverse orientation functionally linked with a promoter. Provided that the promoter and preferably termination and polyadenylation signals are positioned correctly, the strand of the inserted sequence is transcribed which corresponds to the non-coding strand acting as an antisense RNA.

The present invention also relates to ribozymes which cleave specifically the RNA molecules according to the invention and thus are also of use for inhibiting the gene expression. Useful ribozymes may comprise 5'-terminal and 3'-terminal sequences which are complementary to the target RNA, and they can be constructed by a person skilled in the art according to standard methods (see e.g. PCT publication WO 83/23572). The ribozymes according to the invention comprise e.g. ribozymes having the features of group I intron ribozymes (Cech, biotechnology 13 (1995), 323) and "hammerhead" ribozymes (Edgington, Biotechnology 10 (1992), 256).

In one embodiment, the ribozymes according to the invention *per se* are used as drugs. In another embodiment, gene therapy methods are employed for the expression of ribozymes in a target cell *ex vivo* or *in vivo*. The methods of administering the ribozymes or of expressing the ribozymes *in vivo* correspond to the methods described above in connection with the RNA molecules according to the invention.

The isolation and characterization of the human NINTROX gene and in particular the mouse homolog of the NINTROX gene allow to establish an animal model which permits to provide therapies and drugs for the above discussed diseases. Providing the sequence of the NINTROX gene enables both diagnosis (post-natally or pre-natally) and therapy of diseases in which the gene expression is characterized by the lack of NINTROX-RNA or an excess of NINTROX-RNA. However, the therapeutic or diagnostic application is not only limited to diseases, which are accompanied by a dysregulation of the expression of a gene controlled by NINTROX-RNA but the RNA molecules modified in accordance with the above described possibilities also offer the chance of using completely new therapeutic agents.

Therefore, the present invention also relates to drugs which contain the above described RNA molecules, vectors, antibodies, antisense RNAs or ribozymes. These drugs optionally contain additionally a pharmaceutically acceptable carrier. The person skilled in the art is familiar with suitable carriers and the formulation of such drugs. Suitable carriers include e.g. phosphate-buffered common salt solutions, water, emulsions, e.g. oil-in-water emulsions, wetting agents, sterile solutions, etc. The drugs can be administered orally or parenterally. The topical intra-arterial (e.g. directly to the tumor), intramuscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal or intranasal administration belong to the methods for the parenteral administration. A suitable dose is determined by the attending physician and depends on various factors, e.g. on the age, sex, patient's weight, stage of a tumor, kind of administration, etc.

The drug according to the invention is used preferably for preventing or treating diseases which are correlated with a disturbed control of gene expression. The drug according to the invention is used particularly preferably for treating tumoral diseases or diseases of the CNS. In this connection, the drug may be used in gene therapy, the above described methods or vectors being usable for introducing the nucleic acids according to the invention. On the other hand, the RNA molecule according to the invention may be administered directly so as to restore normal expression of the gene in cells which no longer have functional copies of the RNA molecule.

The present invention also relates to a diagnostic composition which contains the RNA molecule according to the invention, to the DNA sequence coding for it or a fragment thereof, to the antibody according to the invention or a fragment thereof, or to the antisense RNA according to the invention or a fragment thereof, or to combinations thereof, optionally together with a suitable analytical reagent. By means of this diagnostic composition the detection may be made as to whether the RNA directly controlling the gene expression, e.g. NINTROX-RNA, is present or, as compared to a control, is available in excessively high or low concentration or with a deviating length. In this connection, the antibody or a fragment thereof is preferably used in the above described assays or the antisense RNA or a fragment thereof as a probe in hybridization experiments. For this purpose, the probe preferably has a length of at least 10, more preferably at least 15, bases. Suitable detection methods based on hybridization are known to the person skilled in the art. Suitable labeling for the probe are also known to the person skilled in the art and they comprise e.g. labeling using radioisotopes, bioluminescence,

chemiluminescence, fluorescence markers, metal chelates, enzymes, etc. This process may use methods known to the person skilled in the art as regards the preparation of whole RNA or poly(A)+RNA from biological samples, the separation of the RNAs on gels separating according to size, e.g. denaturing agarose gels, the production and labeling of the probe and the detection of the hybrids, e.g. via "Northern blot". In this connection, diseases are preferably diagnosed as described above in connection with the drugs according to the invention.

A diagnosis can also be made on a DNA level. In this connection, the intactness of the gene which codes for the RNA which is directly involved in the regulation of gene expression, e.g. NINTROX-RNA, is investigated by the above described nucleic acid molecules (e.g. as regards the availability, length or mutations). For this process it is possible to use methods with which the person skilled in the art is familiar as to the preparation of DNA from biological samples, the restriction digestion of the DNA, the separation of the restriction fragments on gels separating according to size, e.g. agarose gels, the production and labeling of the probe and the detection of hybridization, e.g. via "Southern blot". The above detection can also be carried out via PCR. In this connection, primers are used which flank the coding sequence. Here, amplification products of DNA from the tissue in question, which differ e.g. as regards the length or sequence from the amplification products of DNA from healthy tissue, are of diagnostic significance.

The subject matter of the present invention also relates to a non-human mammal whose NINTROX gene is modified, e.g. by

insertion of a heterologous sequence, in particular a selection marker sequence.

The expression "non-human mammal" comprises any mammal whose NINTROX gene may be modified. Examples of such mammals are mouse, rat, rabbit, horse, cow, sheep, goat, monkey, pig, dog and cat, with mouse being preferred.

The expression "NINTROX gene which is modified" signifies that in the NINTROX gene naturally occurring in a human mammal a deletion of about 1 to 2 kb is carried out by standard methods. If desired, a heterologous sequence, e.g. a construct for mediating antibiotic resistance (e.g. a "neo cassette") can be inserted in this deletion. This method is generally described in Schwartzberg et al., Proc. Natl. Acad. Sci. USA, Vol. 87, pp. 3210-3214, 1990, to which reference is made.

A further subject matter of the present invention relates to cells which are obtained from the above non-human mammal. These cells may be present in any form, e.g. in a primary or long-term culture.

A non-human mammal according to the invention can be provided by common methods. A method is favorable which comprises the steps of:

(a) preparation of a DNA fragment, in particular a vector, containing a modified NINTROX gene, the NINTROX gene having been modified by deletion of a homologous sequence and/or insertion of a heterologous sequence, in particular a selectable marker;

(b) preparation of embryonal stem cells from a non-human mammal (preferably mouse);

(c) transformation of the embryonal stem cells of step (b) with the DNA fragment from step (a), the NINTROX gene in the embryonal stem cells being modified by homologous recombination with the DNA fragment from (a);

(d) culturing the cells from step (c);

(e) selection of the cultured cells from step (d) for the absence of the homologous sequence and/or the presence of the heterologous sequence, in particular the selectable marker,

(f) production of chimeric non-human mammals from the cells from step (e) by injection of these cells in mammalian blastocysts (preferably mouse blastocysts), transfer of the blastocysts in pseudo-pregnant female mammals (preferably mouse) and analyses of the resulting offspring for a modification of the NINTROX gene.

The mechanism of the homologous recombination (cf. R.M. Torres, R. Kühn, Laboratory Protocols for Conditional Gene Targeting, Oxford University Press, 1997) is used in step (c) to transfect embryonal stem cells. The homologous recombination between the DNA sequences present in a chromosome and new, added cloned DNA sequences enables the insertion of a cloned gene in the genome of a living cell in place of the original gene. By this method it is possible to obtain via chimeras animals which are homozygous for the desired gene or the desired gene portion of the desired mutation when embryonal germ cells are used.

The expression "embryonal stem cells" comprises any embryonal stem cells of a non-human mammal which are suitable for the mutation of the NINTROX gene. The embryonal mouse stem cells, in particular cells E14/1 or 129/SV, are preferred.

The term "vector" comprises any vector which by recombination with the DNA of embryonal stem cells enables a modification of the NINTROX gene. The vector preferably has a marker with which it is possible to select for present stem cells in which the desired recombination was made. Such a marker is e.g. the loxP/tkneo cassette which by means of the Cre/loxP system can be removed from the genome again.

In addition, the person skilled in the art knows conditions and materials to carry out steps (a) to (f).

A non-human mammal is provided by the present invention whose NINTROX gene is modified. This modification can be an elimination of the gene expression-regulatory function. By means of such a mammal or cells therefrom the gene expression-controlling function of NINTROX can be investigated selectively. Furthermore, it is possible to find substances, drugs and therapy approaches by which a selective influence can be exerted on the controlling function of NINTROX. Therefore, the present invention furnishes a basis for influencing the most varying diseases. Such diseases are e.g. limitations of the CNS functions which reach as far as mental retardation or the induction of cancer resulting from mistakes made in the control of cell proliferation. Furthermore, it should be possible to investigate in more detail and characterize the part of the hippocampus.

The following clones were deposited with DSMZ, *Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH* [German-type collection of micro-organisms and cell cultures], Mascheroder Weg 1b, D-38124 Braunschweig, on May 4, 1998:

DSM 12153:	*E. coli* JFC-484, partial sequence of the human NINTROX-cDNA
DSM 12154:	*E. coli* JFC-622, partial sequence of the murine NINTROX-cDNA
DSM 12155:	*E. coli* JFC-8D3, sequence of the human genomic NINTROX-DNA
DSM 12156:	*E. coli* JFC-P1-165, sequence of the murine genomic NINTROX-DNA

The figures show:

Figure 1: human sequence of the NINTROX gene

Figure 2: murine sequence of the NINTROX gene

Figure 3: sequence comparison between human (top) and murine (bottom) sequences
Solid bars: sequence-conserved regions (b)

Figure 4: energy diagram of the sequences from figure 3

Figure 5: homology comparison of NINTROX from various species
Figure 5a: partial sequence from hamster
Figure 5b: partial sequence from kangaroo
Figure 5c: partial sequence from macaca
Figure 5d: partial sequence from orangutan
Figure 5e: partial sequence from rat

Figure 5f: partial sequence from chimpanzee

The following example explains the invention:

Example 1: Identification and Characterization of the NINTROX Gene

For the identification of transcribed sequences from the region Xq2-7.3 to Yqter, whole RNA was initially isolated from various pig tissues (kidney, heart, spleen, liver, brain, etc.) and transcribed by means of oligo-dT into first strand cDNA. These complex cDNA samples which represent all of the genes transcribed in the respective tissue were then labeled radioactively and hybridized with the Xq27.3-Xqter-specific cosmid library. The cosmid library was in this connection analyzed in the form of cosmid clones arranged systematically on nylon membranes. Then, the cosmid DNA was isolated by the cosmid clones which had positive hybridization signals with the complex cDNA samples, was digested using EcoRI, separated by gel electrophoresis and transferred to nylon membranes. The restriction fragments which then had a positive hybridization with the complex, radioactively labeled cDNA samples were subsequently isolated and labeled radioactively and used for screening a fetal human cDNA library. By this, positive cDNA clones could be isolated which represented the transcript of the NINTROX gene.

Claims

1. An RNA molecule which can bind to a ligand and comprises the following sequence regions:
   (a) a sequence region maintaining the three-dimensional structure of the RNA molecule; and
   (b) a sequence region for the specific binding of the ligand.

2. The RNA molecule according to claim 1, wherein sequence region (a) comprises the DNA sequence shown in fig. 3 without bars at the margin or a sequence which is related thereto and also permits the maintenance of the three-dimensional structure of the RNA molecule.

3. The RNA molecule according to claim 1 or 2, wherein sequence region (b) comprises the DNA sequence shown in fig. 3 with bars at the margin.

4. The RNA molecule according to any one of claims 1 to 3, wherein the ligand is a DNA molecule or a protein.

5. The RNA molecule according to any one of claims 1 to 4, which additionally contains a poly(A) sequence at the 3' end.

6. The RNA molecule according to any one of claims 1 to 5 for the control of gene expression.

7. The DNA sequence which codes for an RNA molecule according to any one of claims 1 to 6.

8. A gene which comprises the sequence shown in fig. 1 or 2.

9. A vector which comprises the DNA sequence according to claim 7 or the gene according to claim 8.

10. The vector according to claim 9, wherein the vector is a plasmid.

11. The vector according to claim 10, wherein the vector is a viral vector.

12. The vector according to claim 11, which is an RNA virus.

13. The vector according to claim 12, which is a retrovirus.

14. The host cell, containing the vector according to any one of claims 9 to 13.

15. The host cell according to claim 14, wherein the host cell is a mammalian cell.

16. An antibody or a fragment thereof, which bind specifically an RNA molecule according to any one of claims 1 to 6.

17. The antibody according to claim 16, wherein the antibody is a monoclonal antibody.

18. An antisense RNA which binds specifically to an RNA molecule according to any one of claims 1 to 6.

19. A ribozyme which cleaves specifically an RNA molecule according to any one of claims 1 to 6.

20. Use of the RNA molecule according to any one of claims 1 to 6, of the vector according to any one of claims 9 to 13, of the antibody or fragment thereof according to claim 16 or 17, of the antisense RNA according to claim 18 or of the ribozyme according to claim 19 for the production of a pharmaceutical preparation for preventing or treating diseases which are connected with a disturbed control of gene expression.

21. Use of the RNA molecule according to any one of claims 1 to 6, of the DNA sequence according to claim 7 or a fragment thereof, of the antibody or fragment thereof according to claim 16 or 17, or of the antisense RNA according to claim 18 or a fragment thereof for the diagnosis of diseases which are connected with a disturbed control of gene expression.

22. Use according to claim 20 or 21, wherein the disease is a tumoral disease or a disease of the central nervous system.

23. A non-human mammal whose NINTROX gene is modified by deletion of a homologous sequence and/or insertion of a heterologous sequence.

24. The non-human mammal according to claim 23, wherein the heterologous sequence is a selection marker sequence.

25. The non-human mammal according to claim 23 or 24, wherein the selection marker sequence conveys resistance to neomycin.

26. A process for the production of a non-human mammal according to any one of claims 23 to 25, characterized by the following steps:

(a) preparation of a DNA fragment, in particular a vector, containing a modified NINTROX gene, the NINTROX gene having been modified by deletion of a homologous sequence and/or insertion of a heterologous sequence, in particular a selectable marker;

(b) preparation of embryonal stem cells from a non-human mammal (preferably mouse);

(c) transformation of the embryonal stem cells from step (b) with the DNA fragment from step (a), the NINTROX gene in the embryonal stem cells being modified by homologous recombination with the DNA fragment from (a),

(d) culturing the cells from step (c),

(e) selection of the cultured cells from step (d) for the absence of the homologous sequence and/or the presence of the heterologous sequence, in particular the selectable marker,

(f) production of chimeric non-human mammals from the cells from step (e) by injection of these cells in mammalian blastocysts (preferably mouse blastocysts), transfer of the blastocysts into false-pregnant female mammals (preferably mouse) and analysis of the resulting offspring for a change of the NINTROX gene.

Abstract of the Disclosure

The invention relates to modularly constructed RNA molecules which can bind to a ligand and which are characterized by two sequence regions, namely a first sequence region which contributes to the maintenance of the three-dimensional structure of the RNA molecule, and a second sequence region which is responsible for the specific binding of the ligand. These RNA molecules, e.g. the NINTROX RNA, can be used for directly influencing the gene expression. The invention also relates to vectors containing the RNA molecules according to the invention as well as to medicaments and diagnostic compositions which contain said RNA molecules or vectors, to an antibody which specifically recognizes these RNA molecules or antisense RNA binding specifically to these RNA molecules, or to ribozymes cleaving these RNA molecules. In addition, the invention relates to non-human mammals whose NINTROX gene is modified by inserting a heterologous sequence and to cells obtained therefrom.

Human sequence of the non-coding RNA gene (including the putative promoter)

```
   1  CTTAGAGTTT CGTGGCTTCA GGGTGGGAGT AGTTGGAGCA TTGGGGATGT
  51  TTTTCTTACC GACAAGCACA GTCAGGTTGA AGACCTAACC AGGGCCAGAA
 101  GTAGCTTTGC ACTTTTCTAA ACTAGGCTCC TTCAACAAGG CTTGCTGCAG
 151  ATACTACTGA CCAGACAAGC TGTTGACCAG GCACCTCCCC TCCCGCCCAA
 201  ACCTTTCCCC CATGTGGTCG TTAGAGACAG AGCGACAGAG CAGTTGAGAG
 251  GACACTCCCG TTTTCGGTGC CATCAGTGCC CCGTCTACAG CTCCCCCAGC
 301  TCCCCCCACC TCCCCCACTC CCAACCACGT TGGGACAGGG AGGTGTGAGG
 351  CAGGAGAGAC AGTTGGATTC TTTAGAGAAG ATGGATATGA CCAGTGGCTA
 401  TGGCCTGTGC GATCCCACCC GTGGTGGCTC AAGTCTGGCC CCACACCAGC
 451  CCCAATCCAA AACTGGCAAG GACGCTTCAC AGGACAGGAA AGTGGCACCT
 501  GTCTGCTCCA GCTCTGGCAT GGCTAGGAGG GGGGAGTCCC TTGAACTACT
 551  GGGTGTAGAC TGGCCTGAAC CACAGGAGAG GATGGCCCAG GGTGAGGTGG
 601  CATGGTCCAT TCTCAAGGGA CGTCCTCCAA CGGGTGGCGC TAGAGGCCAT
 651  GGAGGCAGTA GGACAAGGTG CAGGCAGGCT GGCCTGGGGT CAGGCCGGGC
 701  AGAGCACAGC GGGGTGAGAG GGATTCCTAA TCACTCAGAG CAGTCTGTGA
 751  CTTAGTGGAC AGGGGAGGGG GCAAAGGGGG AGGAGAAGAA AATGTTCTTC
 801  CAGTTACTTT CCAATTCTCC TTTAGGGACA GCTTAGAATT ATTTGCACTA
 851  TTGAGTCTTC ATGTTCCCAC TTCAAAACAA ACAGATGCTC TGAGAGCAAA
 901  CTGGCTTGAA TTGGTGACAT TTAGTCCCTC AAGCCACCAG ATGTGACAGT
 951  GTTGAGAACT ACCTGGATTT GTATACATAC CTGCGCTTGT TTTAAAGTGG
1001  GCTCAGCACA TAGGGTTCCC ACGAAGCTCC GAAACTCTAA GTGTTTGCTG
1051  CAATTTTATA AGGACTTCCT GATTGGTTTC TCTTCTCCCC TTCCATTTCT
1101  GCCTTTTGTT CATTTCATCC TTTCACTTCT TTCCCTTCCT CCGTCCTCCT
1151  CCTTCCTAGT TCATCCCTTC TCTTCCAGGC AGCCGCGGTG CCCAACCACA
1201  CTTGTCGGCT CCAGTCCCCA GAACTCTGCC TGCCCTTTGT CCTCCTGCTG
1251  CCAGTACCAG CCCCACCCTG TTTTGAGCCC TGAGGAGGCC TTGGGCTCTG
1301  CTGAGTCCAA CCTGGCCTGT CTGTGAAGAG CAAGAGAGCA GCAAGGTCTT
1351  GCTCTCCTAG GTAGCCCCCT CTTCCCTGGT AAGAAAAAGC AAAAGGCATT
1401  TCCCACCCTG AACAACGAGC CTTTTCACCC TTCTACTCTA GAGAAGTGGA
1451  CTGGAGGAGC TGGGCCCGAT TTGGTAGTTG AGGAAAGCAC AGAGGCCTCC
1501  TGTGGCCTGC CAGTCATCGA GTGGCCCAAC AGGGGCTCCA TGCCAGCCGA
1551  CCTTGACCTC ACTCAGAAGT CCAGAGTCTA GCGTAGTGCA GCAGGGCAGT
1601  AGCGGTACCA ATGCAGAACT CCCAAGACCC GAGCTGGGAC CAGTACCTGG
1651  GTCCCCAGCC CTTCCTCTGC TCCCCCTTTT CCCTCGGAGT TCTTCTTGAA
```

Fig. 1

1701 TGGCAATGTT TTGCTTTTGC TCGATGCAGA CAGGGGGCCA GAACACCACA

1751 CATTTCACTG TCTGTCTGGT CCATAGCTGT GGTGTAGGGG CTTAGAGGCA

1801 TGGGCTTGCT GTGGGTTTTT AATTGATCAG TTTTCATGTG GGATCCCATC

1851 TTTTTAACCT CTGTTCAGGA AGTCCTTATC TAGCTGCATA TCTTCATCAT

1901 ATTGGTATAT CCTTTTCTGT GTTTACAGAG ATGTCTCTTA TATCTAAATC

1951 TGTCCAACTG AGAAGTACCT TATCAAAGTA GCAAATGAGA CAGCAGTCTT

2001 ATGCTTCCAG AAACACCCAC AGGCATGTCC CATGTGAGCT GCTGCCATGA

2051 ACTGTCAAGT GTGTGTTGTC TTGTGTATTT CAGTTATTGT CCCTGGCTTC

2101 CTTACTATGG TGTAATCATG AAGGAGTGAA ACATCATAGA AACTGTCTAG

2151 CACTTCCTTG CCAGTCTTTA GTGATCAGGA ACCATAGTTG ACAGTTCCAA

2201 TCAGTAGCTT AAGAAAAAAC CGTGTTTGTC TCTTCTGGAA TGGTTAGAAG

2251 TGAGGGAGTT TGCCCCGTTC TGTTTGTAGA GTCTCATAGT TGGACTTTCT

2301 AGCATATATG TGTCCATTTC CTTATGCTGT AAAAGCAAGT CCTGCAACCA

2351 AACTCCCATC AGCCCAATCC CTGATCCCTG ATCCCCTTCCA CCTGCTCTGC

2401 TGATGACCCC CCCAGCTTCA CTTCTGACTC TTCCCCAGGA AGGGAAGGGG

2451 GGTCAGAAGA GAGGGTGAGT CCTCCAGAAC TCTTCCTCCA AGGACAGAAG

2501 GCTCCTGCCC CCATAGTGGC CTCGAACTCC TGGCACTACC AAAGGACACT

2551 TATCCACGAG AGCGCAGCAT CCGACCAGGT TGTCACTGAG AAGATGTTTA

2601 TTTTGGTCAG TTGGGTTTTT ATGTATTATA CTTAGTCAAA TGTAATGTGG

2651 CTTCTGGAAT CATTGTCCAG AGCTGCTTCC CCGTCACCTG GGCGTCATCT

2701 GGTCCTGGTA AGAGGAGTGC GTGGCCCACC AGGCCCCCCT GTCACCCATG

2751 ACAGTTCATT CAGGGCCGAT GGGGCAGTCG TGGTTGGGAA CACAGCATTT

2801 CAAGCGTCAC TTTATTTCAT TCGGGCCCCA CCTGCAGCTC CCTCAAAGAG

2851 GCAGTTGCCC AGCCTCTTTC CCTTCCAGTT TATTCCAGAG CTGCCAGTGG

2901 GGCCTGAGGC TCCTTAGGGT TTTCTCTCTA TTTCCCCCTT TCTTCCTCAT

2951 TCCCTCGTCT TTCCCAAAGG CATCACGAGT CAGTCGCCTT TCAGCAGGCA

3001 GCCTTGGCGG TTTATCGCCC TGGCAGGCAG GGGCCCTGCA GCTCTCATGC

3051 TGCCCCTGCC TTGGGGTCAG GTTGACAGGA GGTTGGAGGG AAAGCCTTAA

3101 GCTGCAGGAT TCTCACCAGC TGTGTCCGGC CCAGTTTTGG GGTCTGACCT

3151 CAATTTCAAT TTTGTCTGTA CTTGAACATT ATGAAGATGG GGGCCTCTTT

3201 CAGTGAATTT GTGAACAGCA GAATTGACCG ACAGCTTTCC AGTACCCATG

3251 GGGCTAGGTC ATTAAGGCCA CATCCACAGT CTCCCCCACC CTTGTTCCAG

3301 TTGTTAGTTA CTACCTCCTC TCCTGACAAT ACTGTATGTC GTCGAGCTCC

3351 CCCCAGGTCT ACCCCTCCCG GCCCTGCCTG CTGGTGGGCT TGTCATAGCC

3401 AGTGGGATTG CCGGTCTTGA CAGCTCAGTG AGCTGGAGAT ACTTGGTCAC

Fig. 1 (cont'd 1)

3451 AGCCAGGCGC TAGCACAGCT CCCTTCTGTT GATGCTGTAT TCCCATATCA

3501 AAAGGCACAG GGGACACCCA GAAACGCCAC ATCCCCCAAT CCATCAGTGC

3551 CAAACTAGCC AACGGCCCCA GCTTCTCAGC TCGCTGGATG GCGGAAGCTG

3601 CTACTCGTGA GCGCCAGTGC GGGTGCAGAC AATCTTCTGT TGGGTGGCAT

3651 CATTCCAGGC CCGAAGCATG AACAGTGCAC CTGGGACAGG GAGCAGCCCC

3701 AAATTGTCAC CTGCTTCTCT GCCCAGCTTT TCATTGCTGT GACAGTGATG

3751 GCGAAAGAGG GTAATAACCA GACACAAACT GCCAAGTTGG GTGGAGAAAG

3801 GAGTTTCTTT AGCTGACAGA ATCTCTGAAT TTTAAATCAC TTAGTAAGCG

3851 GCTCAAGCCC AGGAGGGAGC AGAGGGATAC GAGCGGAGTC CCCTGCGCGG

3901 GACCATCTGG AATTGGTTTA GCCCAAGTGG AGCCTGACAG CCAGAACTCT

3951 GTGTCCCCCG TCTAACCACA GCTCCTTTTC CAGAGCATTC CAGTCAGGCT

4001 CTCTGGGCTG ACTGGGCCAG GGGAGGTTAC AGGTACCAGT TCTTTAAGAA

4051 GATCTTTGGG CATATACATT TTTAGCCTGT GTCATTGCCC CAAATGGATT

4101 CCTGTTTCAA GTTCACACCT GCAGATTCTA GGACCTGTGT CCTAGACTTC

4151 AGGGAGTCAG CTGTTTCTAG AGTTCCTACC ATGGAGTGGG TCTGGAGGAC

4201 CTGCCCGGTG GGGGGGCAGA GCCCTGCTCC CTCCGGGTCT TCCTACTCTT

4251 CTCTCTGCTC TGACGGGATT TGTTGATTCT CTCCATTTTG GTGTCTTTCT

4301 CTTTTAGATA TTGTATCAAT CTTTAGAAAA GGCATAGTCT ACTTGTTATA

4351 AATCGTTAGG ATACTGCCTC CCCCAGGGTC TAAAATTACA TATTAGAGGG

4401 GAAAAGCTGA ACACTGAAGT CAGTTCTCAA CAATTTAGAA GGAAAACCTA

4451 GAAAACATTT GGCAGAAAAT TACATTTCGA TGTTTTTGAA TGAATACAAG

4501 CAAGCTTTTA CAACAGTGCT GATCTAAAAA TACTTAGCAC TTGGCCTGAG

4551 ATGCCTGGTG AGCATTACAG GCAAGGGGAA TCTGGAGGTA GCCGACCTGA

4601 GGACATGGCT TCTGAACCTG TCTTTTGGGA GTGGTATGGA AGGTGGAGCG

4651 TTCACCAGTG ACCTGGAAGG CCCAGCACCA CCCTCCTTCC CACTCTTCTC

4701 ATCTTGACAG AGCCTGCCCC AGCGCTGACG TGTCAGGAAA ACACCCAGGG

4751 AACTAGGAAG GCACTTCTGC CTGAGGGGCA GCCTGCCTTG CCCACTCCTG

4801 CTCTGCTCGC CTCGGATCAG CTGAGCCTTC TGAGCTGGCC TCTCACTGCC

4851 TCCCCAAGGC CCCCTGCCTG CCCTGTCAGG AGGCAGAAGG AAGCAGGTGT

4901 GAGGGCAGTG CAAGGAGGGA GCACAACCCC CAGCTCCCGC TCCGGGCTCC

4951 GACTTGTGCA CAGGCAGAGC CCAGACCCTG GAGGAAATCC TACCTTTGAA

5001 TTCAAGAACA TTTGGGGAAT TTGGAAATCT CTTTGCCCCC AAACCCCCAT

5051 TCTGTCCTAC CTTTAATCAG GTCCTGCTCA GCAGTGAGAG CAGATGAGGT

5101 GAAAAGGCCA AGAGGTTTGG CTCCTGCCCA CTGATAGCCC CTCTCCCCGC

5151 AGTGTTTGTG TGTCAAGTGG CAAAGCTGTT CTTCCTGGTG ACCCTGATTA

5201 TATCCAGTAA CACATAGACT GTGCGCATAG GCCTGCTTTG TCTCCTCTAT

Fig. 1 (cont'd 2)

```
5251  CCTGGGCTTT TGTTTTGCTT TTTAGTTTTG CTTTTAGTTT TTCTGTCCCT
5301  TTTATTTAAC GCACCGACTA GACACACAAA GCAGTTGAAT TTTTATATAT
5351  ATATCTGTAT ATTGCACAAT TATAAACTCA TTTTGCTTGT GGCTCCACAC
5401  ACACAAAAAA AGACCTGTTA AAATTATACC TGTTGCTTAA TTACAATATT
5451  TCTGATAACC ATAGCATAGG ACAAGGGAAA ATAAAAAAAG AAAAAAAAGA
5501  AAAAAAAACG ACAAATCTGT CTGCTGGTCA CTTCTTCTGT CCAAGCAGAT
5551  TCGTGGTCTT TTCCTCGCTT CTTTCAAGGG CTTTCCTGTG CCAGGTGAAG
5601  GAGGCTCCAG GCAGCACCCA GGTTTTGCAC TCTTGTTTCT CCCGTGCTTG
5651  TGAAAGAGGT CCCAAGGTTC TGGGTGCAGG AGCGCTCCCT TGACCTGCTG
5701  AAGTCCGGAA CGTAGTCGGC ACAGCCTGGT CGCCTTCCAC CTCTGGGAGC
5751  TGGAGTCCAC TGGGGTGGCC TGACTCCCCC AGTCCCCTTC CCGTGACCTG
5801  GTCAGGGTGA GCCCATGTGG AGTCAGCCTC GCAGGCCTCC CTGCCAGTAG
5851  GGTCCGAGTG TGTTTCATCC TTCCCACTCT GTCGAGCCTG GGGGCTGGAG
5901  CGGAGACGGG AGGCCTGGCC TGTCTCGGAA CCTGTGAGCT GCACCAGGTA
5951  GAACGCCAGG GACCCCAGAA TCATGTGCGT CAGTCCAAGG GGTCCCCTCC
6001  AGGAGTAGTG AAGACTCCAG AAATGTCCCT TTCTTCTCCC CCATCCTACG
6051  AGTAATTGCA TTTGCTTTTG TAATTCTTAA TGAGCAATAT CTGCTAGAGA
6101  GTTTAGCTGT AACAGTTCTT TTTGATCATC TTTTTTTAAT AATTAGAAAC
6151  ACCAAAAAAA TCCAGAAACT TGTTCTTCCA AAGCAGAGAG CATTATAATC
6201  ACCAGGGCCA AAAGCTTCCC TCCCTGCTGT CATTGCTTCT TCTGAGGCCT
6251  GAATCCAAAA GAAAAACAGC CATAGGCCCT TTCAGTGGCC GGGCTACCCG
6301  TGAGCCCTTC GGAGGACCAG GGCTGGGGCA GCCTCTGGGC CCACATCCGG
6351  GGCCAGCTCC GGCGTGTGTT CAGTGTTAGC AGTGGGTCAT GATGCTCTTT
6401  CCCACCCAGC CTGGGATAGG GGCAGAGGAG GCGAGGAGGC CGTTGCCGCT
6451  GATGTTTGGC CGTGAACAGG TGGGTGTCTG CGTCGTCCA CGTGCGTGTT
6501  TTCTGACTGA CATGAAATCG ACGCCCGAGT TAGCCTCACC CGGTGACCTC
6551  TAGCCCTGCC CGGATGGAGC GGGGCCCACC CGGTTCAGTG TTTCTGGGGA
6601  GCTGGACAGT GGAGTGCAAA AGGCTTGCAG AACTTGAAGC CTGCTCCTTC
6651  CCTTGCTACC ACGGCCTCCT TTCCGTTTGA TTTGTCACTG CTTCAATCAA
6701  TAACAGCCGC TCCAGAGTCA GTAGTCAATG AATATATGAC CAAATATCAC
6751  CAGGACTGTT ACTCAATGTG TGCCGAGCCC TTGCCCATGC TGGGCTCCCG
6801  TGTATCTGGA CACTGTAACG TGTGCTGTGT TTGCTCCCCT TCCCCTTCCT
6851  TCTTTGCCCT TTACTTGTCT TTCTGGGGTT TTTCTGTTTG GGTTTGGTTT
6901  GGTTTTTATT TCTCCTTTTG TGTTCCAAAC ATGAGGTTCT CTCTACTGGT
6951  CCTCTTAACT GTGGTGTTGA GGCTTATATT TGTGTAATTT TTGGTGGGTG
```

Fig. 1 (cont'd 3)

7001 AAAGGAATTT TGCTAAGTAA ATCTCTTCTG TGTTTGAACT GAAGTCTGTA

7051 TTGTAACTAT GTTTAAAGTA ATTGTTCCAG AGACAAATAT TTCTAGACAC

7101 TTTTTCTTTA CAAACAAAAG CATTCGGAGG GAGGGGGATG GTGACTGAGA

7151 TGAGAGGGGA GAGCTGAACA GATGACCCCT GCCCAGATCA GCCAGAAGCC

7201 ACCCAAAGCA GTGGAGCCCA GGAGTCCCAC TCCAAGCCAG CAAGCCGAAT

7251 AGCTGATGTG TTGCCACTTT CCAAGTCACT GCAAAACCAG GTTTTGTTCC

7301 GCCCAGTGGA TTCTTGTTTT GCTTCCCCTC CCCCCGAGAT TATTACCACC

7351 ATCCCGTGCT TTTAAGGAAA GGCAAGATTG ATGTTTCCTT GAGGGGAGCC

7401 AGGAGGGGAT GTGTGTGTGC AGAGCTGAAG AGCTGGGGAG AATGGGGCTG

7451 GGCCCACCCA AGCAGGAGGC TGGGACGCTC TGCTGTGGGC ACAGGTCAGG

7501 CTAATGTTGG CAGATGCAGC TCTTCCTGGA CAGGCCAGGT GGTGGGCATT

7551 CTCTCTCCAA GGTGTGCCCC GTGGGCATTA CTGTTTAAGA CACTTCCGTC

7601 ACATCCCACC CCATCCTCCA GGGCTCAACA CTGTGACATC TCTATTCCCC

7651 ACCCTCCCCT TCCCAGGGCA ATAAAATGAC CATGGAGGGG GCTTGCACTC

7701 TCTTGGCTGT CACCCGATCG CCAGCAAAAC TTAGATGTGA GAAAACCCCT

7751 TCCCATTCCA TGGCGAAAAC ATCTCCTTAG AAAAGCCATT ACCCTCATTA

7801 GGCATGGTTT TGGGCTCCCA AAACACCTGA CAGCCCCTCC CTCCTCTGAG

7851 AGGCGGAGAG TGCTGACTGT AGTGACCATT GCATGCCGGG TGCAGCATCT

7901 GGAAGAGCTA GGCAGGGTGT CTGCCCCCTC CTGAGTTGAA GTCATGCTCC

7951 CCTGTGCCAG CCCAGAGGCC GAGAGCTATG GACAGCATTG CCAGTAACAC

8001 AGGCCACCCT GTGCAGAAGG GAGCTGGCTC CAGCCTGGAA ACCTGTCTGA

8051 GGTTGGGAGA GGTGCACTTG GGGCACAGGG AGAGGCCGGG ACACACTTAG

8101 CTGGAGATGT CTCTAAAAGC CCTGTATCGT ATTCACCTTC AGTTTTTGTG

8151 TTTTGGGACA ATTACTTTAG AAAATAAGTA GGTCGTTTTA AAAACAAAAA

8201 TTATTGATTG CTTTTTTGTA GTGTTCAGAA AAAAGGTTCT TTGTGTATAG

8251 CCAAATGACT GAAAGCACTG ATATATTTAA AAACAAAAGG CAATTTATTA

8301 AGGAAATTTG TACCATTTCA GTAAACCTGT CTGAATGTAC CTGTATACGT

8351 TTCAAAAACA CCCCCCCCCC ACTGAATCCC TGTAACCTAT TTATTATATA

8401 AAGAGTTTGC CTTATAAATT TA

Fig. 1 (cont'd 4)

Murine sequence of the non-coding RNA gene (including the putative promoter)

```
   1  CTTAGAGTTT CGTGGCTTCG GGGTGGGAGT AGTTGGAGCA TTGGGATGTT
  51  TTTCTTACCG ACAAGCACAG TCAGGTTGAA GACCTAACCA GGGCCAGAAG
 101  TAGCTTTGCA CTTTTCTAAA CTAGGCTCCT TCAACAAGGC TTGCTGCAGA
 151  TACTACTGAC CAGACAAGCT GTTGACCAGG CACTCCCCCC AACAATATCC
 201  TCCCTCTTCC CCCCCCCCAC CCCCGCCCCG TGTGCTCGTT AGGGCAATTG
 251  AAAGGACACT CCCATTTTTG GTGCCATTGA TGCCCTGTCC ATAATAGCTT
 301  CCCTGACTTT TACACCACCC CAACTCCCAA TCTGAAGGAC TGGGAGGTGT
 351  GATGCAGGAG AAACTATGGG ACTCTTGGGA GAAGACTATG GAGTTGGCCA
 401  GTGATTAAGG CCCACTAATT CCAACTGTGG TAGCACAGAT CTGGCTCCAC
 451  ATCAACCCAA TCCAAAACTG ACAAGGATAT TTTGCAAAAA AAGAAAGTGG
 501  CACCTGTCTG ATCCAGCTCT GACATGGCTA GAGGTGAGTC CTAAACTGAT
 551  GGCTTATAAA CTAGCCTGAG CCACAGAAGA GTATGGCCCA GAGTGAAGTG
 601  TCATCATCTG TTCACAAGGC ATGCTCCCCT AGAAGATAAT GCTAAAGAGG
 651  TGCCATGGAG GCAGCAGGAC AAAGTACAGG CAGGCTAGGT GGAGTCAAGC
 701  CAGGCCTAGT GCCACAGAAC AAGAGAGCAG TCTGACTAGT AATTAAGAGG
 751  GAAGAAAGGA AAATATTCTT CCAATTACTT TCCAGTTCTC CTTTAGGGAC
 801  AGCTTAGAAT TATTTGCACT ATTGAGTCTT CATGTTCCCA CTTCAAAACA
 851  AACAGATGCT CTGAAAGCAA ACTGGCTTGA AATGGTGACA CTGTCCCACA
 901  AGCCACCAGA CATGGCAGTG TTCAGAACTA CCTGTATCTG TATATACCTG
 951  CGCTTGTTTT AAAGTGGGCT CAGCACATAG GATTCCCAAG AAGCTCCGAA
1001  ACTCTAAGTG TTTGCTGCAA TTTTATAAGG ACTTCCTGAT TGCTTTCTCT
1051  CTCGTCCTTC CATTTCTTCC TTCCTTCCAT TTCATGCTTT CATTTCTTCC
1101  CCTAGCTTCT AGTTGTTTCT TCTGTTCCAG GCAGCTGCAG TGCTGAACCA
1151  CATGGTTACC TAACAGCAGT CAGCTGCAGC CCTAGGATTC TTCCTGCCCT
1201  TTAACTTCCC ATTGCCAGTG CCAGGTATCA TATTTAACCT TGAGCAAGAG
1251  CTGGGCTCTT TTGAGCCCTC CCTAACCTCT GTGAAGAAGA ACAAGAAGGT
1301  AGGAAGCTCT TGCTCTTGCT AAGAAAAATG TCAAAAGGCT TTCAGACCTT
1351  AAACAATGAG CCTTTTCACC TTTTACTCTA GAAAAGTGGA CTAGAAAATC
1401  TGGGTCACAT TGGGTAGCTG AAGGAGATAC AGAGGCCCCT ATGGCCTGCC
1451  AGAGTCGTTG CATGGCCCAA CAGGGGCTCC ATGCCCACTA CCCTTGACCC
1501  TACTCAGAAA TCTAATGTCA TACTTAGTGT GGGCAGGGGA CCTGTCAGGA
1551  CAGATGCAGA CCTAAGCAGG GAGTGACACC AGGGCCCTTG GCCCTTCTTC
1601  TGACAAACAT ACACATCCCA AGTCTTTTTC TAGTGGAATT CTTAACCTCT
1651  TGCTCACTGG GGACTGGGAA GCATCAGCAC ATCCCATATT TCAAACTCTG
```

Fig. 2

1701 CTCCATAAGT ACAGTGGTGA ATTTTATAGA CTTGACTTTG CTGTGGGGTT
1751 TTAATTGGTC AGTTTTAATT TGGGATCCCA AAGTTTTAAC CTCCATTCAG
1801 GAAGTCCTTA TCTAGCTGCA TATCTTCATC ATATTGGTAT ATCCTTTTCT
1851 GTGTTTACAG AGATGTCTCA TATCTATCGA AATCTGTCTG AGAAGTACCT
1901 TATCAAAGTA GCAAATGAGA CAGCAGTCTT ATGCTTCCAG AAACACCCAC
1951 AGGCACGTCC CATGTGAGCT GCTGCCATGA ACTGTCGAGT GTGTATTGTC
2001 TTGTGTATTT TCGTTAACGT TCCCCAGCTT CCTTCCTGCG GTGTAATCAT
2051 GGAAGAGTGA AACATCATAG AAATCGTCTA GCACTTCCTG GCCAGTCCTT
2101 AGTGATCAGG AACCGTAGTT GACAGTTCCA ATTGATAGCT TAAGATAAAA
2151 CCATGTTTGT CTCTTATGGA ATGGTTAGAA CTAAGTGAGA GATCTTGCCC
2201 CATTCTGTTT GCCGAATCAT AGTTGGACTT TTAGTGTATT TGTATCCATT
2251 TCCTTGTGCT ATAAAAGCAA ACCCTGCAAC CAGCTTTCTG TCAGGCAGTC
2301 CTTTTGCCTG CTCTGCTTTT GATCCTCTTA GTCTTGCTTC TGGTTCCTCC
2351 CTGGAGAGGG AGGAGGGGTC AGAAGAGGAA TTCTGGAGGA TCCAGGATAT
2401 GTCCTTCTGA ACTCCTGCTT CTTCCAGTGA CAAAAGGCCC CTACTGCCCC
2451 ACCCCAACCT GCCCCATGCA CTCCTCTAGG ACACCTTTCC ATACTTTTCA
2501 CAACACCTAG CCAGGTTGAC ACCAAGTTGT TTATTGTGGT CTGCTTGGAA
2551 TTTTACCTGT TAGGCTTACT TAGTCCAATC AAATGGACTC CAAGTTGGGT
2601 ATCCCTCATC TTTGGAAGAC AACCTAGGCT GATTAGATAT TTACTTTTGG
2651 GATTGCAGCA CTTTGGGTGC CGTTTTTCTT TTACTTGGGT TTTATCTGCA
2701 GCTCCCTCAC CACCACCACC ACCCCCCACT TACCTGTATG TAGAACTGAT
2751 TTCAAAACTG CAGGTGGTGG TAACTGCAGC TTCTTAGGGT TTTCTTCACT
2801 TCTTGCTTCT TTCCCCATTC CCTCATCCAC AAATAAGGGC ATCACAAGTC
2851 AGTCTCCTTT AAGCAGGCAG CTTTGGTGGG GTTTTTCCCC TGGAAGCCAG
2901 GGACCCTGTC AGGCTGCCTC TGCCTTGTGG TCAGGTTGAC AGGAGGTTGG
2951 AGGGAAAAGC CTTAAGTCAT GGGATTCTCA CCAGCTGTGT CTGGCTCAGA
3001 CCTGGAATGT GACCTTTATT TTGTTGTATT TGAACATTGT AAAGTGTGGG
3051 TGGTACCTTA AACTGAATAT GTGAAGAATC CAGAAACTGA CCAACAGCTT
3101 TCAGATACCT GGGGCTAGGT CACTAAGGTC ACATCCAGTC TTCCCTACCC
3151 TGTTCTAGTT GTTAGCTACT ACCTCTCCCA GATAGATTGC TGTATATCCT
3201 CCAACTATGA TCATCCTGGC CCAAGCTTGC CTGTTCTTGA GTCTGTCTTA
3251 ACCAGTGGAA CTGCTGCCCT TGGTGTGCAG TGAGTTGAGG ACTCTTGGTC
3301 ACAGCCAGGC TCTAGTAGTA CAGCTCCTTT CTGCTGGTGC TGTATTTCCA
3351 TATCAAAAGG CACAGGGGAG ATCTAGAAAT GCCATCTCCC CCAGTCCATC
3401 AGTGCCAAAC AAGCCCATGA TCCCAGCATG GGTACAGACA ACTCTGTTCA

Fig. 2 (cont'd 1)

09/720215


```
3451  GTGCTATCAC AACAGACTAG AGGCCATGAA CATTGGACGT GGGAACCAGA
3501  GCAACCCGAA TTGCTGCTGC TTTATTCAGC TTTCCGTTGC TCTGACAATG
3551  ATAAAACAAG GCAGTAACTT AAAACAGACT GCCAGGTTTG GCAGAGAAAG
3601  GAAATTCCTT AGCTGACAGC ACCTCTGGAT TTTAAATAGG TTGTAATAAG
3651  TGGCTCAAAC CCATCCAGGA AAAAGCAAAA GGGTTAGAAC TGACCAGATG
3701  AGACCAGCCT GATTTCATGC AGCCCAAATG GAGTCCAGCT GTCTGAACTC
3751  TGCAGCACTT CTCTACTACA GTCTCCTAGA GCATTCCAGC CAGGCTCTTC
3801  AGGCTGAGGA GACATCACAG GTGCCAGTTC TTCAAGAAGA CTTTTGTGCA
3851  TCAGTTCATA GCCTATATCT TTGCCCAAGA TTGTAGATTC AGGTTAACAC
3901  TACAGATTCT AGGGCAGATG ACTGAGACTC AGAAAAAAAG CCCCTGTGGA
3951  CTGTGGTATA GCGAAGTACA AAAACTGAAG GGGGCTAGGG CAGATGCCGC
4001  ATGCCTCATG CCAGAGCCAA GCCCTCTGCT CCATCCACAT CCTTTTCTGG
4051  CTCCTTCTTC CTGCTCTCTG CTTCAGTGAA CCAGCCCCAC TCTGAAGAGA
4101  TTTGTTGATT CTCTCCATTT TTATGTCTTT CTCTTTTAGG TACTATATAG
4151  AAAAGGCTTA GTCTAATTGT TATAAATTGC TAGAATACTG CCTCCCCCAG
4201  GGTCTAAAAA TATATGCTAA AGGGGAAAAC TTGAACACTG AAACCAGTTC
4251  TGAACAATTT AGAAGGAAAA CCTTGAAAAC ATTTAACAAA AAATTATATT
4301  TTAATGTTTA TGAATAAGAG GAGGCTTTTG AAAAAATGTT GATCTATAAA
4351  TACTTACTTT AGGCCTGAGG TGTCTAATGA GTGAACTGAG CAATGGGAAC
4401  TCAAGGCTGA AGCCTCCTGC ATCAGAGGAG GTAGAACCAG GAGCCTCTTG
4451  AGATTTGAGG TGTTTTAGCA TTGGAAAGCC ACTCTTTGGG TAGCTGGCCC
4501  CAGAAACTAC TTCTGACCTT GTCATTTGGA ATGGAGGTTA GTGGTCTGCC
4551  AGATGCCAAA GCTGCATGAG ACCAGCTCTT GGTTTATCAA TTTGAACACT
4601  CAGTAACCTA GAAGGCCCAG CACAAAGTGT CTGCTCTCTT CTTAACTGAG
4651  CCTGCCCCAG CACTACTGCA CAAATTAGGG AGGGTCTACT TCCTACAGAG
4701  CATCCCTCCC TGGGCCCCCT CCCATCCTTT GTACTCTACC TACCTGACCT
4751  TCAGGATCTT GGCACATACG AAATGGCTGT GTAGCAAGCA CTTTGGCATG
4801  CCCTCCTAAA CTTACCCCAG AGCCTCTCCC TGCCTCCTTA AGCCAGTCTG
4851  CCTGTCTTCT GGGGAGGTGT TAGAGCCCAT AGAATGGAGA GGAGAAAGAA
4901  AAGAGGAAGA GGCAGGCAGG TAGTAAAAAG GCTCTGGGAG GAAAGACAGC
4951  CTCCTAGGCT TTGCACAAGC AGGACTCAGC CCCTTGTGGG AACTAAGTGC
5001  CATCTTGGAG TTTAAGAACA TTTGGACAAG TTGCAAATGA CCTTTGCTCC
5051  TTGCTCCTCT CACCTTTTAT GGGGCCCTGC TTAGCACTGA AAGCAAATGC
5101  GCTGAAAAGG CAAAGAGGTT TGGCTCCTGC CCACTGATAG TCCTTTCCCT
5151  GCAGTGTTTG TGTGTCAAGT GGCAAAGCTG TTCTTCCTGG TGACTCTGAT
5201  TAGATCCAGT AACTTAAGAG ATTTGTATGC ATAGGTCTGC TTTGACTCTT
```

Fig. 2 (cont'd 2)

```
5251  CTATTCTGGG CTTTTGATTT GTTTTTCAGT TTTGCTTTTA GTTTTCCTAT
5301  TTTTATTTTA TGCACCAACT AGACACACAA AGCAGTTGAA TTTATATATA
5351  TATATATATA TATATATCTG TATATTTCAC AATTATAAAC TCATTTTGCT
5401  TGTGACGCCA CACACACACA AAAAGAAAAA CCTTTTAAAA TTATACCTGT
5451  TGCTTAATTA CAATATTTCT GATAACCATA GAGTAGGACA AGGGAAAAAA
5501  TTTAAAAAAA AAAAAAAAAA AAGAAAAAAC ACATCTGTCT GCTGGTCACT
5551  TCTTCAATCC AAGCAGATCT GTGATCTTTC CTCGCGTCTT TCAAAGACTT
5601  CCCTGTGCTA AGTGAAGGAA GCTCCAGGCT GCACCCAGGT TTTGTGCTTT
5651  GTTTCTCCTC TGTTGTGAAA GGGGCCCAA GATTCTGGGT ACAGGACAGT
5701  TCATTTCAGC ATGGGGTCAG GAGACAAGAG CACTCCCTTT ACATGCTGAC
5751  GTACAGAACT TAGTGGGAAT AGCCTAGTCC CCACCTCTAG GGATGGGGAG
5801  CTAGCATGCA TGGGGGTGAC CCAACTCCCT CCACCTTTCC CTGGCCAGGA
5851  AGAGCCTGTG TACAGTAAGT CTGACAAGCT TTCCCCAGTT AGCAGGGCTC
5901  AGAGCATTTA AAAACCCTCC AAACTTTGCT GAGTCTAGGG ACTAGAGAGA
5951  AGATAGAAGA TTTGGTCTAT CTCCAAGGTG TGTAAGCTGT ACCAGGTAGA
6001  ATGCCAGGGA CCCCAGAACC ACATCCAACA GCCCAATGGG TCTCCTCCAG
6051  AAAGTAGTGA AGACTCCAGA AACATCCCTT TCTCTTCTCC CTGCTCCCAT
6101  GAGTAACTGC ATTTGCTTTT GTAATCCTTA ATGAGCATTA TCTGCTAAAA
6151  AAAAAAAATT AGCTGTAACA GTTCTTTTTG CAAAAGGATC ATTCTTAAAT
6201  AATTAAAAAC ACCCCCCCCC CAAAAAAAAG TCCAGAACCT TGTTCTTCCA
6251  AAGCAGAGAG CATTATAATC AGGGCCAAAA TCTGTCCCAC ACCTCTACCC
6301  CATCTCCTCA TGATTGCTGC TTCTAAGGCC AGAATACAGC AAAGATATTT
6351  GTAGGCCCTT TGGGTGACTG GGCTACCCTT GGAGCTCTTG GAAGATGGGC
6401  TGGGGAAGCC TCTGAGACCC TATCCTAGGG CCTTGCTCTA GGGAGTAATC
6451  AGTATTAGTA GAGTGTCACA ACATTATTCC CCAGCCGGCA TGAGATGGGG
6501  GCAGAAGAAG CCAAAGGGTT GTCTCCACTG CTACTTACTT GGCCACTGAC
6551  AGGTAGGTGA CCATGTATGT CCATATGCAT GTTTTATGGC TGATGTGAGA
6601  TCAGCACCCA AGTTAGCTTC ACCTGGTGAC CTCTAACCCT GCCTGGATGG
6651  AGCAGGCCAC CTGGTTCAAT GTTTCTGGGC AGCTGGACAA TGGAGTGCAA
6701  AAGGCTTACA GAACTTGAAG CCTTTTCCTT ACTTTGCTAG CACGGCCTCC
6751  TTTTCCATTT GATTTGTCAC TGCTTCAGTC AATAACAGCC GCTCCAGAGT
6801  CAGTAGTTGA TGAATATATG ACCAAATATC ACCAGGACTG TTACTCAACG
6851  TGTGCCGAGC CCTTTCCTTG TGCTGGGCTC CCTGTGTACC TGGACACTGT
6901  AATGTGTGCT GTGTTTGCTC TCCTTCCTCT TCCTTCCTTG CCCTTTCCTT
6951  GTCTTTCTGG GGTTTTTCTG TTGGGTTTGG TTTGGTTTTA TTTTTCCTTT
```

Fig. 2 (cont'd 3)

7001 TGTGTTCCAA ACATGAGGTT TTCTCTACTG GTCCTCTTTA ACTGTGGTGT

7051 TGAGGCTTCT ATTTGTGTAA TTTTTGGTGG GTGAAAGGAA CTTTGCTAAG

7101 TAAATCTCTT CTGTGTTTGA AATGAAGTCT GTATTGTAAC TATGTTTAAA

7151 GTAATTGTTC CAGAGACAAA TGCTTCTAGG TACATTTTCA TTACAAACAA

7201 AGCATTTGAA GGGAGGGAAG TGGTGAATAA GACAAGAGGG GCAATCTGAA

7251 TTGATCCCTG CCCAGATCAG CCAGAAGCTA CCAAAAGTTA AGCACTGGTT

7301 TTCCATTCCA AGTCAAGAGA CTGAAGCTGA TGTTTTGCCA TTTTCAAAGT

7351 CAAAGCAAAA CCAGCTTTTC CACCCAATGG ATTCTTTGCT TCTCCTTCCC

7401 AGATTATTAC TACTGCTGTA ATAATCTAGG AGTGCCAGGA GGGAAAGGAG

7451 TATTAACACA GAGCTGTGCT CACTGAGTAT GGAAAGGCTT GGTCTGAGTT

7501 TTCAGGAGGA TGACCCACTG TGGACATGGG GAGAAGACAG AAGATAAATT

7551 AGCCGCTCCC TGCCTAAGAT ACCTCTTAAT AGATAAGTCA AGGCCATGGA

7601 CATTATTGTC TACAAGGCAT GTTTCAAAGA CATGACCAGT CAGGACACTT

7651 CTGTCATACT CCATGTTGCC CCCTAGTACA CAGTACTAAT CTGATATCTC

7701 TGTTCCCGCC ATGCCTGGGG GATAAAATGA TAGCAGAGAC TCCTTTCCTT

7751 CAATGTGATC TAATTCCCAA CAAAATCTGG GCCTGAGATA CCACCTGTTT

7801 CTATGGCAAA CATCCTCAGT AAAGTGTTAT TCTCATTGCA GATTGTTCCA

7851 GCCTAATGTA AGAGGAACAG AGCAGTGTTC CCTTGGAGCC TCATGTGGAC

7901 AGTTCTACCT GTAGTGACCA GTTGGCTATA GTAGTTATTA GCTGGAACAA

7951 CCAGACAGGG TACATGCCCC CTCCAAAATC CATGTTGTAC TCCCCTCTGC

8001 CAGCCAGGGG GGGTGAGATC TGTAGAATAG TGCAGCCAGT GACAAGCCAC

8051 CTTGTGTTTG TCACCAGCTC AAAAACTCAT CTAAGGTTGG GAGCAGGCAG

8101 ACAAGGCAGA GAGAAAGATC CAGGACAGAC CTAGCTGGGC TGGAGGGGTC

8151 TTGAAAAGCC CTCTGTCGTA TTCACCTTCA GTTTTTGTGC TTTGGGACAA

8201 TTACTTTAGA AAATAAGTAG GTCGTTTTAA AAACAAAATA TTGATTGCTT

8251 TTTTGTAGTG TTCAAAACAA AAGGTTCTTT GTGTATAGCC AAATGACTGA

8301 AAGCACTGAT ATATTTAAAA ACAAAAGGCA ATTTATTAAG GAAATTTGTA

8351 CCATTTCAGT AAACCTGTCT GAATGTACCT GTATACGTTT CAAAAACACA

8401 CCCCACTGAA CCCCTGTAAC CTATTTATTA TATAAAGAGT TTGCCTTATA

8451 AATTTACATA AAAA

Fig. 2 (cont'd 4)

```
Human  1 CTTAGAGTTTCGTGGCTTCAGGGTGGGAGTAGTTGGAGCATTGGGGATGT
Mous   1 ------------------G------------------------.-------

   51 TTTTCTTACCGACAAGCACAGTCAGGTTGAAGACCTAACCAGGGCCAGAA
   50 --------------------------------------------------

  101 GTAGCTTTGCACTTTTCTAAACTAGGCTCCTTCAACAAGGCTTGCTGCAG
  100 --------------------------------------------------

  151 ATACTACTGACCAGACAAGCTGTTGACCAGGCACCTCCCC..........
  150 ----------------------------------TC----CAACAATATC

  191 .........TCCCGCCCAAACCTTTCCCCCATGTGGTCGTTAGAGACAGA
  200 CTCCCTCTTC---C---CC---CCCG----G----C-----.........

  232 GCGACAGAGCAGTTGAGAGGACACTCCCGTTTTCGGTGCCATCAGTGCCC
  241 .....--G---A----A-----------A----T--------TGA-----

  282 CGTC...TACAGCTCCCCCAGCTCCCCCCACCTCCCCCACTCCCAACCAC
  286 T---CATA-T----T---TGA--.TTTA----A----A--------..T-

  329 GTTGGGACAGGGAGGTGTGAGGCAGGAGAGACAGTTGGATTCTTTAGAGA
  333 TGAA----T-----------T--------A--TA-G---C----GG----

  379 AG...ATGGATATGACCAGTGGCTATGGCCTGTGCGATCCCACCCGTGGT
  383 --ACT-----GT--G------AT--A----.CACTA--T---A-T-----

  426 GGCTCAAGTCTGGCCCCACACCAGCCCCAATCCAAAACTGGCAAGGACGC
  432 A--A--GA------T-----T--.A---------------A------TAT

  476 TTCACAGGACAGGAAAGTGGCACCTGTCTGCTCCAGCTCTGGCATGGCTA
  481 --TG--AA-A-A------------------A----------A-------.

  526 GGAGGGGGGAGTCCCTTGAACTACTGG.GTGTAGACTGGCCTGAACCACA
  530 ..--A--T------..-A----GA---CT-A--A---A------G-----

  575 GGAGAGGATGGCCCAGGGTGAGGTGGCATGGTCCATTCTCAAGGGACG.T
  576 -A----T---------A----A---T---CA--TG---A-----C-T-C-

  624 CCTCCAACGGGTGGCGCTAGAG....GCCATGGAGGCAGTAGGACAAGGT
  626 --C-T-GAA-A-AAT----A--AGGT-------------C-------A--

  670 GCAGGCAGGCTGGCCTGGGGTCAGGCCGGGCAGAGCACAGCGGGGTGAGA
  676 A----------A-.G---A----A---A---CT--TG-CA-A-AACA--.

  720 GGGATTCCTAATCACTCAGAGCAGTCTGTGACTTAGTGGACAGGGGAGGG
  724 .................-----------ACTAG--A-.............

  770 GGCAAAGGGGGAGGAGAAGAAAATGTTCTTCCAGTTACTTTCCAATTCTC
  744 ...T---A----A--A-G------A--------A----------G-----

  820 CTTTAGGGACAGCTTAGAATTATTTGCACTATTGAGTCTTCATGTTCCCA
  791 --------------------------------------------------

  870 CTTCAAAACAAACAGATGCTCTGAGAGCAAACTGGCTTGAATTGGTGACA
  841 ------------------------A----------------A--------

  920 TTTAGTCCCTCAAGCCACCAGATGTGACAGTGTTGAGAACTACCTGGATT
  891 C-..-----A------------CA--G-------C-----------T--C

  970 TGTATATATACCTGCGCTTGTTTTAAAGTGGGCTCAGCACATAGGGTTCC
  939 --..-------------------------------------------A----

 1020 CACGAAGCTCCGAAACTCTAAGTGTTTGCTGCAATTTTATAAGGACTTCC
  987 --A-----------------------------------------------

 1070 TGATTGGTTTCTCTTCTCCCCTTCCATTTCTGCCTTTTGTTCATTTCATC
 1037 ------C-------CTCGT------------T----CCT-C--------G

 1120 CTTTCACTTCTTTCCCTTCCTCCGTCCTCCTCCTTCCTAGTTCATCCCTT
 1087 ------T-----C----AG-..............-T------.G-TT---

 1170 CTCTTCCAGGCAGCCGCGGTGCCCAACC.............ACACTTGTC
 1122 --G-----------T--A----TG----ACATGGTTACCTA---GCA---

 1207 GGCTCCAGTCCCCAGAACTCTGCCTGCCCTTTGTCCTCCTGCTGCCAGTA
 1172 A---G---.---T--G-T---T-----------AA-T---CAT-------G

 1257 CCAGCCCCACCCTGTTTTGAGCCCTGAGGAGGCCTTGGGCTCTGCTGAGT
 1221 ----..GT-T-A-A---..-A--T----C-A-AGC--------TT----C

 1307 CCAACCTGGCCTGTCTG.TGAAGAGCAAGAGAGCAGCAAGGTCTTGCTCT
 1267 --TC---AA---C-G--AA-----A-----AG-T--G---C---------

 1356 CCTAGGTAGCCCCCTCTTCCCTGGTAAGAAAAA..GCAAAAGGCATTTCC
 1317 .........................--C----------TGT---------.----A

 1404 CACCCTGAACAACGAGCCTTTTCACCCTTCTACTCTAGAGAAGTGGACTG
 1345 G---T-A-----T------------.----T----------A---------A

 1454 GAGGAGCTGGGCCCGATTTGGTAGTTGAGGAAAGCACAGAGGCCTCCTGT
 1394 --AA-T-----T-AC---G-----C---A-G-GAT---------.---A-

 1504 GGCCTGCC..AGTCATCGAGTGGCCCAACAGGGGCTCCATGCCAGCCGAC
 1443 --------AG----G-T-CA------------------------CA-TAC-

 1552 CTTGACCTCACTCAGAAGTCCAGAGTCTAGCGTAGTGCAGCAGGGCAGTA
 1493 -------CT--------A--T-AT---ATA-T-----....T------..

 1602 GCGGTACCAATGCAGAACTCCCAAGACCCGAGCTGGGACCAGTACCTGGG
 1537 ..--G---TG-CAG--CAGATGC-----TA---A----GTGAC---A---

 1652 TCCCCAGCCCTTCCTCTGCTCCCCCTTTTCCCTCGGAGTTCTTCTTGAAT
 1585 C--TTG-------T----ACAAA-A-ACA-ATC-CA---CT--T-CT-G-

 1702 GGCAATGTTTTGCTTTTGCTCGATGCAGACAGG...GGGCCAGAACACCA
 1635 --A-T-C--AAC--C------AC--GG---T--GAA-CAT---C---T-C

 1749 CACATTTCACTGTCTGTCTGGTCCATAGCTGTGGTGTAGGGGCTTAGAGG
 1685 --T------....AAC----C------AG-ACA---GT-AATT---T--A

 1799 CATGGGCTTGCTGTGGGTTTTTAATTGATCAGTTTTCATGTGGGATCCCA
 1731 -T--ACT----------G---------G--------A--T----------

 1849 TCTTTTTAACCTCTGTTCAGGAAGTCCTTATCTAGCTGCATATCTTCATC
 1781 AAG----------CA-----------------------------------

 1899 ATATTGGTATATCCTTTTCTGTGTTTACAGAGATGTCTCTTA..TATCTA
 1831 ---------------------------------------A--TC----G-

 1947 AATCTGTCCAACTGAGAAGTACCTTATCAAAGTAGCAAATGAGACAGCAG
 1881 -------....---------------------------------------

 1997 TCTTATGCTTCCAGAAACACCCACAGGCATGTCCCATGTGAGCTGCTGCC
 1927 ------------------------------C-------------------

 2047 ATGAACTGTCAAGTGTGTGTTGTCTTGTGTATTTCAGTTATTG.TCCCTG
 1977 -----------G-------A----------------TC----AC-T----CA

 2096 GCTTCCTTACTATGGTGTAATCATGAAGGAGTGAAACATCATAGAAACTG
 2027 --------C--GC------------G-A--------------------TC-

 2146 TCTAGCACTTCCTTGCCAGTCTTTAGTGATCAGGAACCATAGTTGACAGT
 2077 -------------G-------C----------------G-----------

 2196 TCCAATCAGTAGCTTAAGAAAAAACCGTGTTTGTCTCTTCTGGAATGGTT
 2127 ------TGA-----------A------A------------A----------

 2246 AG....AAGTGAGGGAGTTTGCCCCGTTCTGTTTGTAGAGTCTCATAGTT
 2177 --AACT-------A--TC-------A---------CC--..A--------

 2292 GGACTTTCTAGCATATATGTGTCCATTTCCTTATGCTGTAAAAGCAAGTC
 2225 -------.---TG---T---A-----------G----A---------AC-

 2342 CTGCAACCAAACTCCCATCAGCCCAATCCCTGATCCCTGATCCCTTCCAC
 2274 ---------GCT-T-TG----G-...............-GT----TTG-

 2392 CTGCTCTGCTGATGACCCCCCCAGCTTCACTTCTGACTCTTCCCCAGGAA
 2308 ----------TT---T--T-TT--TC-TG------GT--C----TG-AG-

 2442 GGGAAGGGGGGTCAGAAGAG................AGGGTGAGTCCTCC
 2358 ----G-A-------------GAATTCTGGAGGATCC---A-AT-----T-

 2476 AGAACT....CTTCCTCCAAGGACAGAAGGCTCCTGCCCCCATAGTGGCC
 2408 T-----CCTG----T----GT----A-----C---A-TG--CC-....--

 2522 TCGAACT...CCTGGCACTACCAAAGGACACTTATCCA.CGAGAGCGCAG
 2454 C-A-C--GCC--AT-----C-TCT-------C-T----TACTTTT-A--A

 2568 CATCCGACCAGGTTGTCACTGAGAAGATGTTTATTTTGGTCAG.TTGGGT
 2504 --C-TAG--------A---...C---T--------G-----T-C----AA

 2617 TTTTATGTATTA....TACTTAGTCAAATGTAATGTGGCTTCTGGAATCA
 2551 -----CC-G---GGCT---------C---CA----...............

 2663 TTGTCCAGAGCTGCTTCCCCGTCACCTGGGCGTCATCTGGTCCTGGTAAG
 2586 .............-AC---AA--TGGG-ATCCC------.....-T-G---

 2713 AGGAGTGCGTGGCCCACCAGGCCCCCCTGTCACCCATGACAGTTCATTCA
 2619 -.............................................--AC-T

 2763 GGGCCGATGGGGCAGTCGTGGTTGGGAACACAGCATTTCAAGCGTC.ACT
 2626 A---T---TA-AT-T-TACTT------TTG-----C--TGG-T-C-GTT-

 2812 TTATTTCATTCGGGCCCCACCTGCAGCTCCCTCAAAGAGGCAGTTGCCCA
 2676 --C---T-C-T---TTTT-T--------------..-CCAC--CCAC-A-C

 2862 GCCTCTTTCCCT.....TCCAGTTTATTCCAGAGCTGCCAGTGGGG...C
 2724 C--CAC--A---GTATG-AG-AC-G---T--A-A----AG----T-GTAA

 2904 CTGAGGCTCCTTAGGGTTTTCTCTCTATTTCCCCCTTTCTTCCTCATTCC
 2774 ---CA---T-----------..---TC-C---TTG---CT----C------

 2954 CTCGTCTTTCCCAAA....GGCATCACGAGTCAGTCGCCTTTCAGCAGGC
 2822 ---A--...-A----TAAG--------A--------T-----A-------

 3000 AGCCTTGG.CGGTTTATCGCCCTGGCAGGCAGGGGCCCTGCAGCTCTCAT
 2869 ---T----TG--G--T-TC------A--C-----A-----......---G

 3049 GCTGCCCCTGCCTTGGGGTCAGGTTGACAGGAGGTTGGAGGG.AAAGCCT
 2913 ------T--------T---------------------------A-------

 3098 TAAGCTGCAGGATTCTCACCAGCTGTGTCCGGCCCAGTTTTGGGGTCTGA
 2963 ----TCATG--------------------T---T---ACC---AA-G---

 3148 CCTCAATTTCAATTTTGTCTGTACTTGAACATTATGAA..GATGGGGGCC
 3013 --......-TT-------....T.........G.A..GT.TG..T..TA

 3196 TCTTTCAGTGAATTTGTGAACA..GCAG.AATTGACCGACAGCTTTCCAG
 3056 C---AA-C-----A------G-ATC---A--C-----A---------AGA

 3243 TACCCATGGGGCTAGGTCATTAAGGCCACATCCACAGTCTCCCCCACCCT
 3106 ----..-------------C-----T------..------T---T----.

 3293 TGTTCCAGTTGTTAGTTACTACCTCCTCTCCTGACAATACTGTATGTCGT
 3151 -----T---------C---------TC-CAGAT-G-T-G------A--C-

 3343 CGAGCTCCCCCCAGGTCTACCCCTCCCGGCCCTGCCTGCTGGTGGGCTTG
 3201 -C-A--.....AT-A--AT--TGG---AAG-T------T-CT--A-TC--

 3393 TCATAGCCAGTGGGATTGCCGGTCTTGACAGCTCAGTGAGCTGGAGATAC
 3246 --T--A-------A-C---T-CC----.GT-TG-------T--AG--CT-

 3443 TTGGTCACAGCCAGGCGC...TAGCACAGCTCCCTTCTGTTGATGCTGTA
 3295 -----------------T-TAG---T--------T-----C--G-------

 3490 TTCCCATATCAAAAGGCACAGGGGACACCCAGAAACGCCACATCCCCCAA
 3345 --T----------------------G-T-T-----T----TC-------G

 3540 TCCATCAGTGCCAAACTAGCCAACGGCCCCAGCTTCTCAGCTCGCTGGAT
 3395 -----------------A----C-T-AT------A-..............

 3590 GGCGGAAGCTGCTACTCGTGAGCGCCAGTGCGGGTGCAGACAATCTTCTG
 3430 ...............................----A-------..C----

 3640 TTGGGTGGCATCATTCCAGGCCCGAAG.CATGAACAGTGCACCTGGGACA
 3447 --CA---CT----CAA---A-TA--G-C--------T--G--G-----AC
```

Fig. 3 (1)

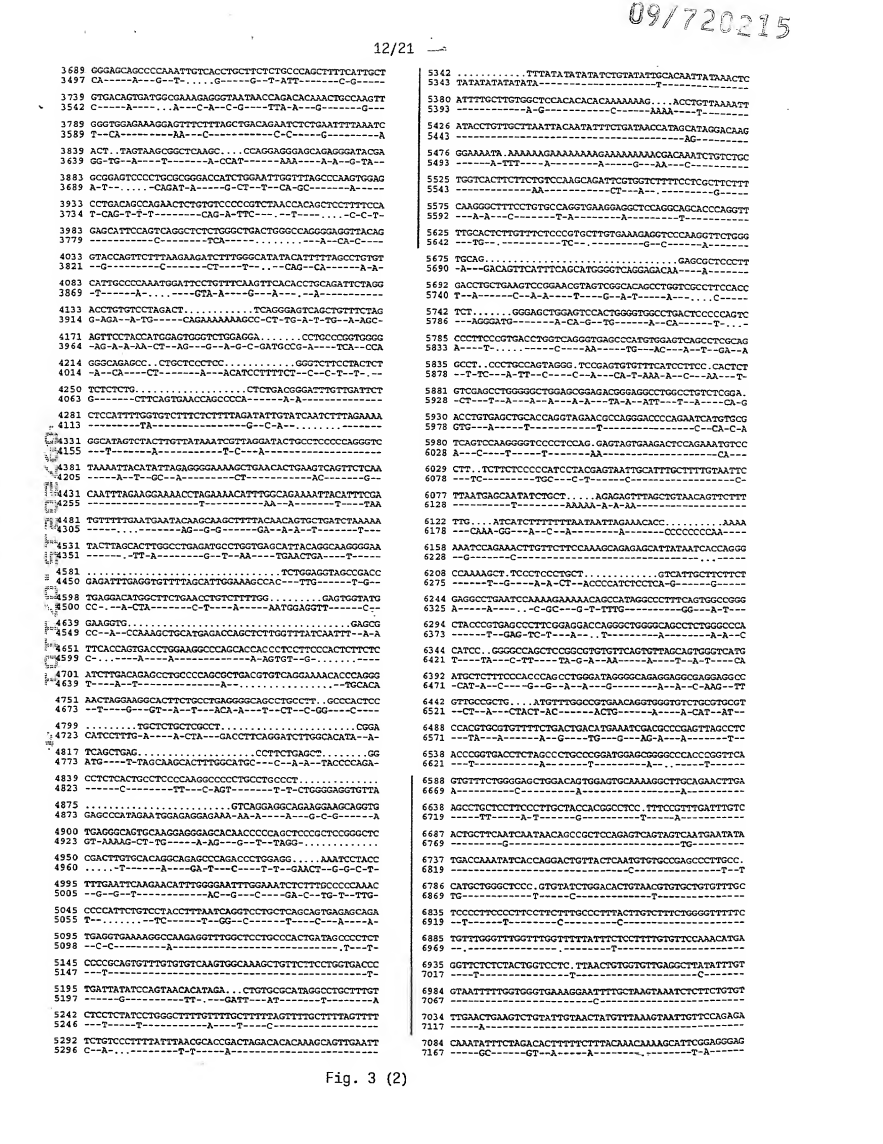

Fig. 3 (2)

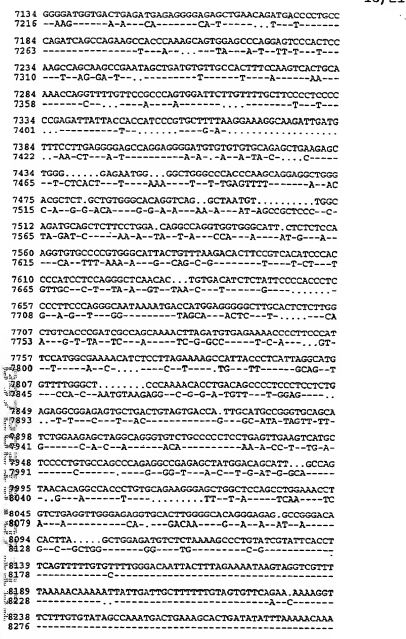

Fig. 3 (3)

dashed line: putative promoter

full line: sequence-conserved high-energy sequence

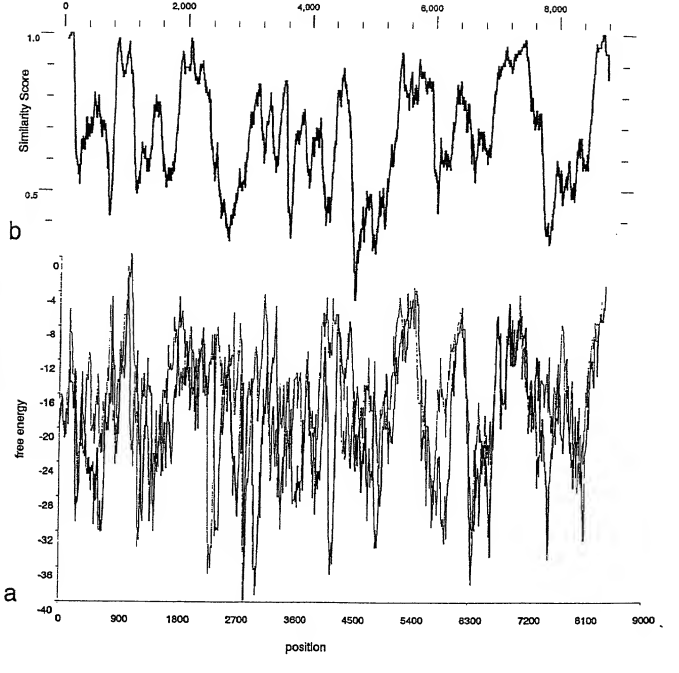

Fig. 4

black similarity 100 window
blue hinlex 10 HUMAN

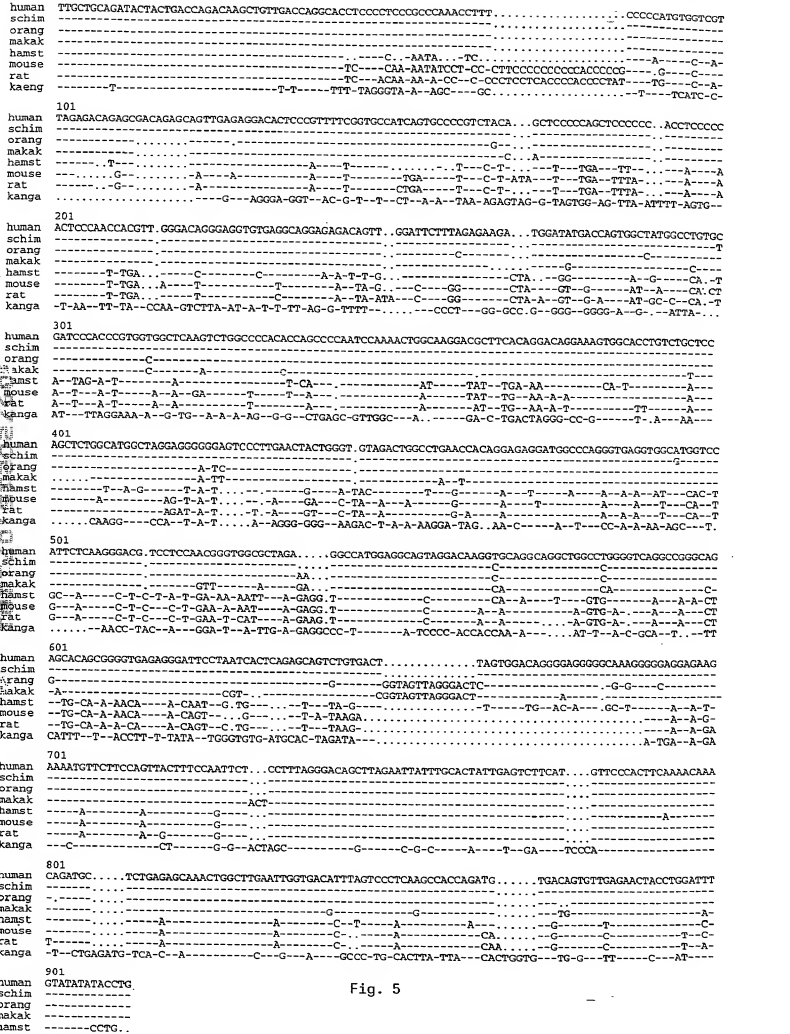

Fig. 5

Partial sequence of the non-coding RNA gene from hamster

```
  1  TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCCCCCCA
 51  ATACTCCCCC AATGTGCTCA TTAGAGATAG CAGTTGAGAG GACACTCCCA
101  TTTTTGGTGC CCTGTCCATA GCTTCCCTGA CTCTTCCACC ACCCCAACTC
151  CCAATCTGAG GGACCGGGAG GTGCGAGGCA GGAAAAATAT TGGATTCTTT
201  AGAGAAGACT AGAGGTGACC AGTGACTGTG GCCCAGTAAT TAGAACTGTG
251  GTGGCACAAG TCTGGCCCCA CATCCACCCA ATCCAAAACT GATAAGGATA
301  TTTTGAAAAA CAGGAAAGCA GTACCTGTCT GATCCAGCTC TGGTATAGGT
351  AGGAGTGAGT CCTGAACTGC TGGATTACAG ACTGGCTTGA GCCACAGAAG
401  ATGATGGACC AGAGTAAAGT ATCATCACCT GCTCACAAGG CATGCTTCAC
451  TAGAGAATAA TTCTAAAGAG GTGCCATGGA GGCAGCAGGA CAAGGCACAA
501  GCAGTCTGGG TGGGGGTCAA GCCAGACCTA GTGCCACAGA ACAAGAGAGC
551  AATCTGTGAC TAGTAGTTAG GGACTTTGTG GATGGGACAA GGGGCATGGG
601  GGAAGAAATG AAAATATTCT TCCAATTACT TTCCAGTTCT CCTTTAGGGA
651  CAGCTTAGAA TTATTTGCAC TATTGAGTCT TCATGTTCCC ACTTAAAAAC
701  AAACAGATGC TCTGAAAGCA AACTGGCTTG AAATGGTGAC ACTTTGTCCC
751  ACAAGCCACC AAATGTGGCA GTGTTTAGAA CTACCTGGAT CTGTATATAC
801  CTG
```

Fig. 5a

Partial sequence of the non-coding RNA gene from kangaroo

```
  1  TTGCTGCATA TACTACTGAC CAGACAAGCT GTTTATCAGG CTTTTTAGGG
 51  TACACCAGCA CCTGCCCTCC ATTCATCCCT GTTGGGAGAG GGATGGTGTA
101  CTGGTTGTCA CTAGAGACCT AACAGAGTAG GGTTAGTGGG AGCTTACATT
151  TTCAGTGCCA TTAACATTCT AGTCCAAGGT CTTAAATTAT TATGTTGAGG
201  GGTTTTTTTT CCCCTGAGGG GGCCGGGGGG TGGGGGGAGG GTTGATTAGA
251  TTCCTTAGGA AAGAGGGTTG AGACAGACAG CAGAGCACTG AGCAGTTGGC
301  ACTAAAGGAG ACCTTGACTA GGGGCCAGGT GGCATCATCT AATCCCAAGG
351  GGCTCCAAGT GAGTATTAGG GTGGGGGAAG ACATTATAGA AGGAATAGAA
401  ACAGGATAGC TCAGCCTAAA GAAGAGCGGT TAAAACCCTA CCCACCAGGA
451  GTTGACTTGA AAGAGGCCCC TATGGAGGAA TCCCCAACCA CCAAAAGCAA
501  TCTTGAGCTG CAGCTGCTTC ATTTAGTGGA CCTTGTGTAT ATCTGGGTGT
551  GTATGCACAT AGATAGACAG TGAGAAAGAA AACTGTTCTT CCAGTTCTTT
601  TCCAGTGCTA CTAGCTTAGG GACAGGTTAG AACTGTCTGC ACAATTGTGT
651  GATCATTCCC ATTCCCACTT CAAAACAAAC TGACTGAGAT GTTCAACAGA
701  AAACTGGCTT CAATGGGTAA CATGCCCTTG CCACTTACTT AAGACACTGG
751  TGTGATGGGG TTTTGAACTC CCTATATTTG TAGGTATCTG
```

Fig. 5b

Partial sequence of the non-coding RNA gene from makaka

```
  1  TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCTCCCCT
 51  CCCGCCCAAA CCTTTCCCCC ATGTGGTCGT TAGAGACAGA GCAGTTGAGA
101  GGACACTCCC GTTTTCGGTG CCATCAGTGC CCCGTCTACC ACTCCCCCAG
151  CTCCCCCCAC CTCCCCCACT CCCAACCACG TTGGGACAGG GAGGTGTGAG
201  GCAGGAGAGA CAGTTGGATT CTTTAGAGAT GGATGTGACC AGTGGCTATG
251  GCCCGTGCGA TCCCACCCGT GGCGGCTCAA ATCTGGCCCC ACCCCAGCCC
301  CAATCCAAAA CTGGCAAGGA CGCTTCACAG GACAGGAAAG TGGCACCTGT
351  CTGTTCCGGC ATGGCTAGGA GGGAGTTGTC CCTTGAACTA CTGGGTGTAG
401  ACTGGCCTAA ATCACAGGAG AGGATGGCCC AGGGTGAGGT GGCATGGTCC
451  ATTCTCAAGG GACGTCCTCC AGTTGGTGGC ACTAGAGAGG CCATGGAGGC
501  AGTAGGACAA GGCACAGGCA GGCTGGCCCA GGGTCAGGCC GGGCCGAACA
551  CAGCGGGGTG AGAGGGATTC CTCGTCTCAG AGCAGTCTGT GACCGGTAGT
601  TAGGGACTTA GTGGACAGGG AAGGGGCAAA GGGGGAGGAG AAGAAAATGT
651  TCTTCCAGTT ACTTTCCAAT TCTACTCCTT TAGGGACAGC TTAGAATTAT
701  TTGCACTATT GAGTCTTCAT GTTCCCACTT CAAAACAAAC AGATGCTCTG
751  AGAGCAAACT GGCTTGAATT GGTGACGTTT AGTCCCTCAG GCCACCAGAT
801  GTGATGGTGT TGAGAACTAC CTGGATATGT ATATATACCT G
```

Fig. 5c

Partial sequence of the non-coding RNA gene from orangutan

```
  1  TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCTCCCCT
 51  CCCGCCCAAA CCTTTCCCCC ATGTGGTCGT TAGAGACAGA GCAGTTGAGA
101  GGACACTCCC GTTTTCGGTG CCATCAGTGC CCCGTCTGCA GCTCCCCCAG
151  CTCCCCCCAC CTCCCCCACT CCCAACCACG TTGGGACAGG GAGGTGTGAG
201  GCAGGAGAGA CAGTTGGATT CTTTCGAGAA GATGGATATG ACCAGTGGCC
251  ATGGCCTGTG CGATCCCACC CGTGGCGGCT CAAGTCTGGC CCCACACCAG
301  CCCCAATCCA AAACTGGCAA GGACGCTTCA CAGGACAGGA AAGTGGCACC
351  TGTCTGCTCC AGCTCTGGCA TGGCTAGGAG GGAGTCGTCC CTTGAACTAC
401  TGGGTGTAGA CTGGCCTGAA CCACAGGAGA GGATGGCCCA GGGTGAGGTG
451  GCATGGTCCA TTCTCAAGGG ACGTCCTCCA ACGGGTGGCG CTAGAAAGGC
501  CATGGAGGCA GTAGGACAAG GCGCAGGCAG GCTGGCCCGG GGTCAGGCCG
551  GGCAGGGCAC AGCGGGGTGA GAGGGATTCC TAATCACTCA GAGCAGTGTG
601  TGACTGGTAG TTAGGGACTC AGTGGACAGG GGAGGGGCGA GGGGGCAGGA
651  GAAGAAAATG TTCTTCCAGT TACTTTCCAA TTCTCCTTTA GGGACAGCTT
701  AGAATTATTT GCACTATTGA GTCTTCATGT TCCCACTTCA AAACAAACGA
751  TGCTCTGAGA GCAAACTGGC TTGAATTGGT GACATTTAGT CCCTCAAGCC
801  ACCAGATGTG AGTGTTGAGA ACTACCTGGA TTTGTATATA TACCTG
```

Fig. 5d

Partial sequence of the non-coding RNA gene from rat

```
  1  TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACTCCCCAC
 51  AACAACAACC CCCTCCCTCC TCACCCCACC CCTATCCCCT GTGTGCTCAT
101  TAGAGAGGGC AATTGAGAGG ACACTCCCAT TTTTGGTGCC ACTGATGCCC
151  TGTCCATAGC TTCCCTGACT TTTACACCAC CCCAACTCCC AATCTGAGGG
201  ACTGGGAGGT GTGACGCAGG AGAAACTATA TAGGACTCTT GGGAGAAGAC
251  TATAGAGTTG GCAAGTGATT GCGCCCCAGT AATTCCAACT GTGGTAGCAC
301  AAGTCTGGCT CCACACCAAC CCAATCCAAA ACTGACAAGG ACATTTTGCA
351  AAAAATGAAA GTGGCATTTG TCTGATCCAG CTCTGGCATG GCTAGAGATG
401  AGTCTTAAAC TGTTGGCTTA TAAACTGGCC TGAGCAACAG AAGAGGATGG
451  CCCAGAGTAA AGTGTCATCA TCTGTTCACA AGGCATGCTC CCCTAGAAGT
501  TCATGCTAAA GAAGTGCCAT GGAGGCAGCA GGACAAAGTA CAGGCTAGGT
551  GGAGTCAAGC CAGGCCTAGT GCCACAGAGC AAGAGAGCAG TCTCTGACTA
601  GTAGTTAAGG GGGAAGAAAG AAAAATATTC TTCCAATTGC TTTCCAGTTC
651  TCCTTTAGGG ACAGCTTAGA ATTATTTGCA CTATTGAGTC TTCATGTTCC
701  CACTTCAAAA CAAATAGATG CTCTGAAAGC AAACTGGCTT GAAATGGTGA
751  CACTGTCCCA CAAGCCACCA GACAATGGCA GTGTTCAGAA CTACCTGTAT
801  ATGTATATAC CTG
```

Fig. 5e

Partial sequence of the non-coding RNA gene from chimpanzee

```
  1  TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCTCCCCT
 51  CCCGCCCAAA CCTTTCCCCC ATGTGGTCGT TAGAGACAGA GCGACAGAGC
101  AGTTGAGAGG ACACTCCCGT TTTCGGTGCC ATCAGTGCCC CGTCTACAGC
151  TCCCCCAGCT CCCCCCACCT CCCCCACTCC CAACCACGTT GGGACAGGGA
201  GGTGTGAGGC AGGAGAGACA GTTGGATTCT TTAGAGAAGA TGGATATGAC
251  CAGTGGCTAT GGCCTGTGTG ATCCCACCCG TGGTGGCTCA AGTCTGGCCC
301  CACACCAGCC CCAATCCAAA ACTGGCAAGG ACGCTTCACA GGACAGGAAA
351  GTGGCACCTG TCTGCTCCAG CTCTGGCATG GCTAGGAGGG GGGAGTCCCT
401  TGAACTACTG GGTGTAGACT GGCCTGAACC ACAGGAGAGG ATGGCCCAGG
451  GTGAGGTGGC GTGGTCCATT CTCAAGGGAC GTCCTCCAAC GGGTGGCGCT
501  AGAGGCCATG GAGGCAGTAG GACAAGGCGC AGGCAGGCTG GCCCGGGGTC
551  AGGCCGGGCA GAGCACAGCG GGGTGAGAGG GATTCCTAAT CACTCAGAGC
601  AGTCTGTGAC TTAGTGGACA GGGGAGGGGG CAAAGGGGGA GGAGAAGAAA
651  ATGTTCTTCC AGTTACTTTC CAATTCTCCT TTAGGGACAG CTTAGAATTA
701  TTTGCACTAT TGAGTCTTCA TGTTCCCACT TCAAAACAAA CAGATGCTCT
751  GAGAGCAAAC TGGCTTGAAT TGGTGACATT TAGTCCCTCA AGCCACCAGA
801  TGTGACAGTG TTGAGAACTA CCTGGATTTG TATATATACC TG
```

Fig. 5f

COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY (Includes Reference to Provisional and PCT International Applications)	Attorney's Docket No. 012627-019

As a below named inventor, I hereby declare that:
My residence, post office address and citizenship are as stated below next to my name;
I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

MODULARLY CONSTRUCTED RNA MOLECULES HAVING TWO SEQUENCE REGION TYPES

the specification of which (check only one item below):

☐ is attached hereto.

☐ was filed as United States application
Number _____
on _____
and was amended
on _____ (if applicable).

☒ was filed as PCT international application
Number PCT/DE99/01867
on 25 June 1999
and was amended
on _____ (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 (a)-(e) of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

PRIOR FOREIGN/PCT APPLICATION(S) AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. §119:

COUNTRY (if PCT, indicate "PCT")	APPLICATION NUMBER	DATE OF FILING (day, month, year)	PRIORITY CLAIMED UNDER 35 U.S.C. §119
DE	198 28 624.4	26 June 1998	X Yes _No
			_Yes _No
			_Yes _No
			_Yes _No
			_Yes _No

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below.

_____ _____
(Application Number) (Filing Date)

_____ _____
(Application Number) (Filing Date)

(01/01)

COMBINED DECLARATION FOR PATENT APPLICATION AND POWER OF ATTORNEY (CONT'D)
(Includes Reference to Provisional and PCT International Applications)

Attorney's Docket No. 012627-019

I hereby claim the benefit under Title 35, United States Code, §120 of any United States applications(s) or PCT international application(s) designating the United States of America that is/are listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in that/those prior application(s) in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose to the Office all information known to me to be material to the patentability as defined in Title 37, Code of Federal Regulations §1.56, which became available between the filing date of the prior application(s) and the national or PCT international filing date of this application:

PRIOR U.S. APPLICATIONS OR PCT INTERNATIONAL APPLICATIONS DESIGNATING THE U.S. FOR BENEFIT UNDER 35 U.S.C. §120:

U.S. APPLICATIONS		STATUS *(check one)*		
U.S. APPLICATION NUMBER	U.S. FILING DATE	PATENTED	PENDING	ABANDONED

PCT APPLICATIONS DESIGNATING THE U.S.					
PCT APPLICATION NO.	PCT FILING DATE	U.S. APPLICATION NUMBERS ASSIGNED (if any)			

I hereby appoint the following attorneys and agent(s) to prosecute said application and to transact all business in the Patent and Trademark Office connected therewith and to file, prosecute and to transact all business in connection with international applications directed to said invention:

William L. Mathis	17,337	Eric H. Weisblatt	30,505	Bruce T. Wieder	33,815
Robert S. Swecker	19,885	James W. Peterson	26,057	Todd R. Walters	34,040
Platon N. Mandros	22,124	Teresa Stanek Rea	30,427	Ronni S. Jillions	31,979
Benton S. Duffett, Jr.	22,030	Robert E. Krebs	25,885	Harold R. Brown III	36,341
Norman H. Stepno	22,716	William C. Rowland	30,888	Allen R. Baum	36,086
Ronald L. Grudziecki	24,970	T. Gene Dillahunty	25,423	Steven M. duBois	35,023
Frederick G. Michaud, Jr.	26,003	Patrick C. Keane	32,858	Brian P. O'Shaughnessy	32,747
Alan E. Kopecki	25,813	B. Jefferson Boggs, Jr.	32,344	Kenneth B. Leffler	36,075
Regis E. Slutter	26,999	William H. Benz	25,952	Fred W. Hathaway	32,236
Samuel C. Miller, III	27,360	Peter K. Skiff	31,917	Wendi L. Weinstein	34,456
Robert G. Mukai	28,531	Richard J. McGrath	29,195	Mary Ann Dillahunty	34,576
George A. Hovanec, Jr.	28,223	Matthew L. Schneider	32,814		
James A. LaBarre	28,632	Michael G. Savage	32,596		
E. Joseph Gess	28,510	Gerald F. Swiss	30,113		
R. Danny Huntington	27,903	Charles F. Wieland III	33,096		

21839

and: _____

Address all correspondence to:

21839

Teresa Stanek Rea
BURNS, DOANE, SWECKER & MATHIS, L.L.P.
P.O. Box 1404
Alexandria, Virginia 22313-1404

Address all telephone calls to: Teresa Stanek Rea at (703) 836-6620.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

(01/01)

FULL NAME OF SOLE OR FIRST INVENTOR	SIGNATURE	DATE
Annemarie POUSTKA		15.3.01
RESIDENCE		CITIZENSHIP
Werderstrasse 36, D-69120 Heidelberg, DE DEX		DE
POST OFFICE ADDRESS		
Werderstrasse 36, D-69120 Heidelberg, DE		
FULL NAME OF SECOND JOINT INVENTOR, IF ANY	SIGNATURE	DATE
Johannes COY		15.3.01
RESIDENCE		CITIZENSHIP
In den schwarzen Garten 1, D-63762 Gross-Ostheim, DE		DE
POST OFFICE ADDRESS		
In den schwarzen Garten 1, D-63762 Gross-Ostheim, DE		
FULL NAME OF THIRD JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		CITIZENSHIP
POST OFFICE ADDRESS		
FULL NAME OF FOURTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		CITIZENSHIP
POST OFFICE ADDRESS		
FULL NAME OF FIFTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		CITIZENSHIP
POST OFFICE ADDRESS		
FULL NAME OF SIXTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		CITIZENSHIP
POST OFFICE ADDRESS		
FULL NAME OF SEVENTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		CITIZENSHIP
POST OFFICE ADDRESS		
FULL NAME OF EIGHTH JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE		CITIZENSHIP
POST OFFICE ADDRESS		

09/720215

Patent

Attorney's Docket No. 012627-019

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of	
Annemarie Poustka et al	Group Art Unit: Not yet assigned
Serial No.: 09/720,215	Examiner: Not yet assigned
Filed: December 22, 2000	**ATTENTION: BOX SEQUENCE**
For: Modularly Constructed RNA Molecules Having Two Sequence Region Types	

DECLARATION PURSUANT TO 37 C.F.R. §§1.821-1.825

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

I, Teresa Stanek Rea, declare as follows:

1. That the content of the paper and computer readable copies of the Sequence Listing, submitted in accordance with 37 C.F.R. §1.821(c) and (e), respectively, are the same in compliance with §1.821(f).

2. That the submission, filed in accordance with 37 C.F.R. §1.821(g)[or (h)], herein does not include new matter [or go beyond the disclosure in the international application].

3. That the substitute copy of the computer readable form, submitted in accordance with 37 C.F.R. §1.825(d), is identical to that originally filed.

I hereby declare that all statements made herein of my own knowledge are true and that all statements were made on information and belief and are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

7/11/01
Date

Teresa Stanek Rea
Registration No. 30,427

SEQUENCE LISTING

<110> Poustka, Annemarie
Coy, Johannes

<120> Modularly Constructed RNA Molecules Having Two Sequence Region Types

<130> 012627-019

<140> US 09/720,215
<141> 2000-12-22

<150> PCT/DE99/01867
<151> 1999-06-25

<150> DE 198 28 624.4
<151> 1998-06-26

<160> 8

<170> PatentIn version 3.0

<210> 1
<211> 8422
<212> DNA
<213> Human

<400> 1
cttagagttt cgtggcttca gggtgggagt agttggagca ttggggatgt ttttcttacc 60

gacaagcaca gtcaggttga agacctaacc agggccagaa gtagctttgc acttttctaa 120

actaggctcc ttcaacaagg cttgctgcag atactactga ccagacaagc tgttgaccag 180

gcacctcccc tcccgcccaa acctttcccc catgtggtcg ttagagacag agcgacagag 240

cagttgagag gacactcccg ttttcggtgc catcagtgcc ccgtctacag ctcccccagc 300

tccccccacc tcccccactc ccaaccacgt tgggacaggg aggtgtgagg caggagagac 360

agttggattc tttagagaag atggatatga ccagtggcta tggcctgtgc gatcccaccc 420

gtggtggctc aagtctggcc ccacaccagc cccaatccaa aactggcaag gacgcttcac 480

aggacaggaa agtggcacct gtctgctcca gctctggcat ggctaggagg ggggagtccc 540

ttgaactact gggtgtagac tggcctgaac cacaggagag gatggcccag ggtgaggtgg 600

catggtccat tctcaaggga cgtcctccaa cgggtggcgc tagaggccat ggaggcagta 660

ggacaaggtg caggcaggct ggcctggggt caggccgggc agagcacagc ggggtgagag 720

ggattcctaa tcactcagag cagtctgtga cttagtggac aggggagggg gcaaaggggg 780

aggagaagaa aatgttcttc cagttacttt ccaattctcc tttagggaca gcttagaatt 840

atttgcacta ttgagtcttc atgttcccac ttcaaaacaa acagatgctc tgagagcaaa 900

ctggcttgaa ttggtgacat ttagtccctc aagccaccag atgtgacagt gttgagaact 960

acctggattt gtatatatac ctgcgcttgt tttaaagtgg gctcagcaca tagggttccc 1020

acgaagctcc gaaactctaa gtgtttgctg caatttttata aggacttcct gattggtttc 1080

tcttctcccc ttccatttct gccttttgtt catttcatcc tttcacttct ttcccttcct 1140

ccgtcctcct ccttcctagt tcatcccttc tcttccaggc agccgcggtg cccaaccaca 1200

cttgtcggct ccagtcccca gaactctgcc tgccctttgt cctcctgctg ccagtaccag 1260

ccccaccctg ttttgagccc tgaggaggcc ttgggctctg ctgagtccaa cctggcctgt 1320

ctgtgaagag caagagagca gcaaggtctt gctctcctag gtagcccccct cttccctggt 1380

aagaaaaagc aaaaggcatt tcccaccctg aacaacgagc cttttcaccc ttctactcta 1440

gagaagtgga ctggaggagc tgggcccgat ttggtagttg aggaaagcac agaggcctcc 1500

tgtggcctgc cagtcatcga gtggcccaac aggggctcca tgccagccga ccttgacctc 1560

actcagaagt ccagagtcta gcgtagtgca gcagggcagt agcggtacca atgcagaact 1620

cccaagaccc gagctgggac cagtacctgg gtccccagcc cttcctctgc tccccctttt 1680

ccctcggagt tcttcttgaa tggcaatgtt ttgcttttgc tcgatgcaga cagggggcca 1740

gaacaccaca catttcactg tctgtctggt ccatagctgt ggtgtagggg cttagaggca 1800

tgggcttgct gtgggttttt aattgatcag ttttcatgtg ggatcccatc tttttaacct 1860

ctgttcagga agtccttatc tagctgcata tcttcatcat attggtatat ccttttctgt 1920

gtttacagag atgtctctta tatctaaatc tgtccaactg agaagtacct tatcaaagta 1980

gcaaatgaga cagcagtctt atgcttccag aaacacccac aggcatgtcc catgtgagct 2040

gctgccatga actgtcaagt gtgtgttgtc ttgtgtattt cagttattgt ccctggcttc 2100

cttactatgg tgtaatcatg aaggagtgaa acatcataga aactgtctag cacttccttg 2160

ccagtcttta gtgatcagga accatagttg acagttccaa tcagtagctt aagaaaaaac 2220

cgtgtttgtc tcttctggaa tggttagaag tgagggagtt tgccccgttc tgtttgtaga 2280

gtctcatagt tggactttct agcatatatg tgtccatttc cttatgctgt aaaagcaagt 2340

cctgcaacca aactcccatc agcccaatcc ctgatccctg atcccttcca cctgctctgc 2400

tgatgacccc cccagcttca cttctgactc ttccccagga agggaagggg ggtcagaaga 2460

gagggtgagt cctccagaac tcttcctcca aggacagaag gctcctgccc ccatagtggc 2520

ctcgaactcc tggcactacc aaaggacact tatccacgag agcgcagcat ccgaccaggt 2580

tgtcactgag aagatgttta ttttggtcag ttgggttttt atgtattata cttagtcaaa 2640
tgtaatgtgg cttctggaat cattgtccag agctgcttcc ccgtcacctg ggcgtcatct 2700
ggtcctggta agaggagtgc gtggcccacc aggccccccc gtcacccatg acagttcatt 2760
cagggccgat ggggcagtcg tggttgggaa cacagcattt caagcgtcac tttatttcat 2820
tcgggcccca cctgcagctc cctcaaagag gcagttgccc agcctctttc ccttccagtt 2880
tattccagag ctgccagtgg ggcctgaggc tccttagggt tttctctcta tttccccctt 2940
tcttcctcat tccctcgtct ttcccaaagg catcacgagt cagtcgcctt tcagcaggca 3000
gccttggcgg tttatcgccc tggcaggcag gggccctgca gctctcatgc tgcccctgcc 3060
ttggggtcag gttgacagga ggttggaggg aaagccttaa gctgcaggat tctcaccagc 3120
tgtgtccggc ccagttttgg ggtctgacct caatttcaat tttgtctgta cttgaacatt 3180
atgaagatgg gggcctcttt cagtgaattt gtgaacagca gaattgaccg acagctttcc 3240
agtacccatg gggctaggtc attaaggcca catccacagt ctcccccacc cttgttccag 3300
ttgttagtta ctacctcctc tcctgacaat actgtatgtc gtcgagctcc ccccaggtct 3360
acccctcccg gccctgcctg ctggtgggct tgtcatagcc agtgggattg ccggtcttga 3420
cagctcagtg agctggagat acttggtcac agccaggcgc tagcacagct cccttctgtt 3480
gatgctgtat tcccatatca aaaggcacag gggacaccca gaaacgccac atcccccaat 3540
ccatcagtgc caaactagcc aacggcccca gcttctcagc tcgctggatg gcggaagctg 3600
ctactcgtga gcgccagtgc gggtgcagac aatcttctgt tgggtggcat cattccaggc 3660
ccgaagcatg aacagtgcac ctgggacagg gagcagcccc aaattgtcac ctgcttctct 3720
gcccagcttt tcattgctgt gacagtgatg gcgaaagagg gtaataacca gacacaaact 3780
gccaagttgg gtggagaaag gagtttcttt agctgacaga atctctgaat tttaaatcac 3840
ttagtaagcg gctcaagccc aggagggagc agagggatac gagcggagtc ccctgcgcgg 3900
gaccatctgg aattggttta gcccaagtgg agcctgacag ccagaactct gtgtcccccg 3960
tctaaccaca gctccttttc cagagcattc cagtcaggct ctctgggctg actgggccag 4020
gggaggttac aggtaccagt tctttaagaa gatctttggg catatacatt tttagcctgt 4080
gtcattgccc caaatggatt cctgtttcaa gttcacacct gcagattcta ggacctgtgt 4140
cctagacttc agggagtcag ctgtttctag agttcctacc atggagtggg tctggaggac 4200
ctgcccggtg gggggcaga gccctgctcc ctccgggtct tcctactctt ctctctgctc 4260
tgacgggatt tgttgattct ctccattttg gtgtctttct cttttagata ttgtatcaat 4320

ctttagaaaa ggcatagtct acttgttata aatcgttagg atactgcctc ccccagggtc 4380
taaaattaca tattagaggg gaaaagctga acactgaagt cagttctcaa caatttagaa 4440
ggaaaaccta gaaaacattt ggcagaaaat tacatttcga tgtttttgaa tgaatacaag 4500
caagctttta caacagtgct gatctaaaaa tacttagcac ttggcctgag atgcctggtg 4560
agcattacag gcaaggggaa tctggaggta gccgacctga ggacatggct tctgaacctg 4620
tcttttggga gtggtatgga aggtggagcg ttcaccagtg acctggaagg cccagcacca 4680
ccctccttcc cactcttctc atcttgacag agcctgcccc agcgctgacg tgtcaggaaa 4740
acacccaggg aactaggaag gcacttctgc ctgaggggca gcctgccttg cccactcctg 4800
ctctgctcgc ctcggatcag ctgagccttc tgagctggcc tctcactgcc tccccaaggc 4860
cccctgcctg ccctgtcagg aggcagaagg aagcaggtgt gagggcagtg caaggaggga 4920
gcacaacccc cagctcccgc tccgggctcc gacttgtgca caggcagagc ccagaccctg 4980
gaggaaatcc tacctttgaa ttcaagaaca tttggggaat ttggaaatct ctttgccccc 5040
aaacccccat tctgtcctac ctttaatcag gtcctgctca gcagtgagag cagatgaggt 5100
gaaaaggcca agaggtttgg ctcctgccca ctgatagccc ctctccccgc agtgtttgtg 5160
tgtcaagtgg caaagctgtt cttcctggtg accctgatta tatccagtaa cacatagact 5220
gtgcgcatag gcctgctttg tctcctctat cctgggcttt tgttttgctt tttagttttg 5280
cttttagttt ttctgtccct tttatttaac gcaccgacta gacacacaaa gcagttgaat 5340
ttttatatat atatctgtat attgcacaat tataaactca ttttgcttgt ggctccacac 5400
acacaaaaaa agacctgtta aaattatacc tgttgcttaa ttacaatatt tctgataacc 5460
atagcatagg acaagggaaa ataaaaaaag aaaaaaaaga aaaaaaaacg acaaatctgt 5520
ctgctggtca cttcttctgt ccaagcagat tcgtggtctt ttcctcgctt ctttcaaggg 5580
ctttcctgtg ccaggtgaag gaggctccag gcagcaccca ggttttgcac tcttgtttct 5640
cccgtgcttg tgaaagaggt cccaaggttc tgggtgcagg agcgctccct tgacctgctg 5700
aagtccggaa cgtagtcggc acagcctggt cgccttccac ctctgggagc tggagtccac 5760
tggggtggcc tgactccccc agtccccttc ccgtgacctg gtcagggtga gcccatgtgg 5820
agtcagcctc gcaggcctcc ctgccagtag ggtccgagtg tgtttcatcc ttcccactct 5880
gtcgagcctg ggggctggag cggagacggg aggcctggcc tgtctcggaa cctgtgagct 5940
gcaccaggta gaacgccagg gaccccagaa tcatgtgcgt cagtccaagg ggtcccctcc 6000
aggagtagtg aagactccag aaatgtccct ttcttctccc ccatcctacg agtaattgca 6060

tttgcttttg taattcttaa tgagcaatat ctgctagaga gtttagctgt aacagttctt 6120
tttgatcatc ttttttttaat aattagaaac accaaaaaaa tccagaaact tgttcttcca 6180
aagcagagag cattataatc accagggcca aaagcttccc tccctgctgt cattgcttct 6240
tctgaggcct gaatccaaaa gaaaaacagc cataggccct ttcagtggcc gggctacccg 6300
tgagcccttc ggaggaccag ggctggggca gcctctgggc ccacatccgg ggccagctcc 6360
ggcgtgtgtt cagtgttagc agtgggtcat gatgctcttt cccacccagc ctgggatagg 6420
ggcagaggag gcgaggaggc cgttgccgct gatgtttggc cgtgaacagg tgggtgtctg 6480
cgtgcgtcca cgtgcgtgtt ttctgactga catgaaatcg acgcccgagt tagcctcacc 6540
cggtgacctc tagccctgcc cggatggagc ggggcccacc cggttcagtg tttctgggga 6600
gctggacagt ggagtgcaaa aggcttgcag aacttgaagc ctgctccttc ccttgctacc 6660
acggcctcct ttccgtttga tttgtcactg cttcaatcaa taacagccgc tccagagtca 6720
gtagtcaatg aatatatgac caaatatcac caggactgtt actcaatgtg tgccgagccc 6780
ttgcccatgc tgggctcccg tgtatctgga cactgtaacg tgtgctgtgt ttgctcccct 6840
tccccttcct tctttgccct ttacttgtct ttctggggtt tttctgtttg ggtttggttt 6900
ggtttttatt tctccttttg tgttccaaac atgaggttct ctctactggt cctcttaact 6960
gtggtgttga ggcttatatt tgtgtaattt ttggtgggtg aaaggaattt tgctaagtaa 7020
atctcttctg tgtttgaact gaagtctgta ttgtaactat gtttaaagta attgttccag 7080
agacaaatat ttctagacac tttttcttta caaacaaaag cattcggagg gaggggatg 7140
gtgactgaga tgagagggga gagctgaaca gatgacccct gcccagatca gccagaagcc 7200
acccaaagca gtggagccca ggagtcccac tccaagccag caagccgaat agctgatgtg 7260
ttgccacttt ccaagtcact gcaaaaccag gttttgttcc gcccagtgga ttcttgtttt 7320
gcttcccctc ccccgagat tattaccacc atcccgtgct tttaaggaaa ggcaagattg 7380
atgtttcctt gaggggagcc aggaggggat gtgtgtgtgc agagctgaag agctggggag 7440
aatggggctg ggcccaccca agcaggaggc tgggacgctc tgctgtgggc acaggtcagg 7500
ctaatgttgg cagatgcagc tcttcctgga caggccaggt ggtgggcatt ctctctccaa 7560
ggtgtgcccc gtgggcatta ctgtttaaga cacttccgtc acatcccacc ccatcctcca 7620
gggctcaaca ctgtgacatc tctattcccc accctcccct tcccagggca ataaaatgac 7680
catggagggg gcttgcactc tcttggctgt cacccgatcg ccagcaaaac ttagatgtga 7740
gaaaacccct tcccattcca tggcgaaaac atctccttag aaaagccatt accctcatta 7800

ggcatggttt tgggctccca aaacacctga cagcccctcc ctcctctgag aggcggagag 7860

tgctgactgt agtgaccatt gcatgccggg tgcagcatct ggaagagcta ggcagggtgt 7920

ctgccccctc ctgagttgaa gtcatgctcc cctgtgccag cccagaggcc gagagctatg 7980

gacagcattg ccagtaacac aggccaccct gtgcagaagg gagctggctc cagcctggaa 8040

acctgtctga ggttgggaga ggtgcacttg gggcacaggg agaggccggg acacacttag 8100

ctggagatgt ctctaaaagc cctgtatcgt attcaccttc agtttttgtg ttttgggaca 8160

attactttag aaaataagta ggtcgtttta aaaacaaaaa ttattgattg ctttttttgta 8220

gtgttcagaa aaaaggttct ttgtgtatag ccaaatgact gaaagcactg atatatttaa 8280

aaacaaaagg caatttatta aggaaatttg taccatttca gtaaacctgt ctgaatgtac 8340

ctgtatacgt ttcaaaaaca ccccccccc actgaatccc tgtaacctat ttattatata 8400

aagagtttgc cttataaatt ta 8422

<210> 2
<211> 8464
<212> DNA
<213> Murine

<400> 2
cttagagttt cgtggcttcg gggtgggagt agttggagca ttgggatgtt tttcttaccg 60

acaagcacag tcaggttgaa gacctaacca gggccagaag tagctttgca cttttctaaa 120

ctaggctcct tcaacaaggc ttgctgcaga tactactgac cagacaagct gttgaccagg 180

cactcccccc aacaatatcc tccctcttcc ccccccccac ccccgccccg tgtgctcgtt 240

agggcaattg aaaggacact cccatttttg gtgccattga tgccctgtcc ataatagctt 300

ccctgacttt tacaccaccc caactcccaa tctgaaggac tgggaggtgt gatgcaggag 360

aaactatggg actcttggga gaagactatg gagttggcca gtgattaagg cccactaatt 420

ccaactgtgg tagcacagat ctggctccac atcaacccaa tccaaaactg acaaggatat 480

tttgcaaaaa aagaaagtgg cacctgtctg atccagctct gacatggcta gaggtgagtc 540

ctaaactgat ggcttataaa ctagcctgag ccacagaaga gtatggccca gagtgaagtg 600

tcatcatctg ttcacaaggc atgctcccct agaagataat gctaaagagg tgccatggag 660

gcagcaggac aaagtacagg caggctaggt ggagtcaagc caggcctagt gccacagaac 720

aagagagcag tctgactagt aattaagagg gaagaaagga aaatattctt ccaattactt 780

tccagttctc ctttagggac agcttagaat tatttgcact attgagtctt catgttccca 840

cttcaaaaca aacagatgct ctgaaagcaa actggcttga aatggtgaca ctgtcccaca 900
agccaccaga catggcagtg ttcagaacta cctgtatctg tatatacctg cgcttgtttt 960
aaagtgggct cagcacatag gattcccaag aagctccgaa actctaagtg tttgctgcaa 1020
ttttataagg acttcctgat tgctttctct ctcgtccttc catttcttcc ttccttccat 1080
ttcatgcttt catttcttcc cctagcttct agttgtttct tctgttccag gcagctgcag 1140
tgctgaacca catggttacc taacagcagt cagctgcagc cctaggattc ttcctgccct 1200
ttaacttccc attgccagtg ccaggtatca tatttaacct tgagcaagag ctgggctctt 1260
ttgagccctc cctaacctct gtgaagaaga acaagaaggt aggaagctct tgctcttgct 1320
aagaaaaatg tcaaaaggct ttcagacctt aaacaatgag ccttttcacc ttttactcta 1380
gaaaagtgga ctagaaaatc tgggtcacat tgggtagctg aaggagatac agaggcccct 1440
atggcctgcc agagtcgttg catggcccaa caggggctcc atgcccacta cccttgaccc 1500
tactcagaaa tctaatgtca tacttagtgt gggcagggga cctgtcagga cagatgcaga 1560
cctaagcagg gagtgacacc agggcccttg gcccttcttc tgacaaacat acacatccca 1620
agtctttttc tagtggaatt cttaacctct tgctcactgg ggactgggaa gcatcagcac 1680
atcccatatt tcaaactctg ctccataagt acagtggtga attttataga cttgactttg 1740
ctgtggggtt ttaattggtc agttttaatt tgggatccca aagttttaac ctccattcag 1800
gaagtcctta tctagctgca tatcttcatc atattggtat atccttttct gtgtttacag 1860
agatgtctca tatctatcga aatctgtctg agaagtacct tatcaaagta gcaaatgaga 1920
cagcagtctt atgcttccag aaacacccac aggcacgtcc catgtgagct gctgccatga 1980
actgtcgagt gtgtattgtc ttgtgtattt tcgttaacgt tccccagctt ccttcctgcg 2040
gtgtaatcat ggaagagtga aacatcatag aaatcgtcta gcacttcctg gccagtcctt 2100
agtgatcagg aaccgtagtt gacagttcca attgatagct taagataaaa ccatgtttgt 2160
ctcttatgga atggttagaa ctaagtgaga gatcttgccc cattctgttt gccgaatcat 2220
agttggactt ttagtgtatt tgtatccatt tccttgtgct ataaaagcaa accctgcaac 2280
cagctttctg tcaggcagtc cttttgcctg ctctgctttt gatcctctta gtcttgcttc 2340
tggttcctcc ctggagaggg aggaggggtc agaagaggaa ttctggagga tccaggatat 2400
gtccttctga actcctgctt cttccagtga caaaaggccc ctactgcccc accccaacct 2460
gccccatgca ctcctctagg acacctttcc atacttttca caacacctag ccaggttgac 2520
accaagttgt ttattgtggt ctgcttggaa ttttacctgt taggcttact tagtccaatc 2580

aaatggactc caagttgggt atccctcatc tttggaagac aacctaggct gattagatat 2640
ttacttttgg gattgcagca ctttgggtgc cgtttttctt ttacttgggt tttatctgca 2700
gctccctcac caccaccacc acccccccact tacctgtatg tagaactgat ttcaaaactg 2760
caggtggtgg taactgcagc ttcttagggt tttcttcact tcttgctttct ttccccattc 2820
cctcatccac aaataagggc atcacaagtc agtctccttt aagcaggcag ctttggtggg 2880
gtttttcccc tggaagccag ggaccctgtc aggctgcctc tgccttgtgg tcaggttgac 2940
aggaggttgg agggaaaagc cttaagtcat gggattctca ccagctgtgt ctggctcaga 3000
cctggaatgt gacctttatt ttgttgtatt tgaacattgt aaagtgtggg tggtacctta 3060
aactgaatat gtgaagaatc cagaaactga ccaacagctt tcagatacct ggggctaggt 3120
cactaaggtc acatccagtc ttccctaccc tgttctagtt gttagctact acctctccca 3180
gatagattgc tgtatatcct ccaactatga tcatcctggc ccaagcttgc ctgttcttga 3240
gtctgtctta accagtggaa ctgctgccct tggtgtgcag tgagttgagg actcttggtc 3300
acagccaggc tctagtagta cagctccttt ctgctggtgc tgtatttcca tatcaaaagg 3360
cacaggggag atctagaaat gccatctccc ccagtccatc agtgccaaac aagcccatga 3420
tcccagcatg ggtacagaca actctgttca gtgctatcac aacagactag aggccatgaa 3480
cattggacgt gggaaccaga gcaacccgaa ttgctgctgc tttattcagc tttccgttgc 3540
tctgacaatg ataaaacaag gcagtaactt aaaacagact gccaggtttg gcagagaaag 3600
gaaattcctt agctgacagc acctctggat tttaaatagg ttgtaataag tggctcaaac 3660
ccatccagga aaaagcaaaa gggttagaac tgaccagatg agaccagcct gatttcatgc 3720
agcccaaatg gagtccagct gtctgaactc tgcagcactt ctctactaca gtctcctaga 3780
gcattccagc caggctcttc aggctgagga gacatcacag gtgccagttc ttcaagaaga 3840
cttttgtgca tcagttcata gcctatatct ttgcccaaga ttgtagattc aggttaacac 3900
tacagattct agggcagatg actgagactc agaaaaaaag cccctgtgga ctgtggtata 3960
gcgaagtaca aaaactgaag ggggctaggg cagatgccgc atgcctcatg ccagagccaa 4020
gccctctgct ccatccacat ccttttctgg ctccttcttc ctgctctctg cttcagtgaa 4080
ccagccccac tctgaagaga tttgttgatt ctctccattt ttatgtcttt ctcttttagg 4140
tactatatag aaaaggctta gtctaattgt tataaattgc tagaatactg cctcccccag 4200
ggtctaaaaa tatatgctaa aggggaaaac ttgaacactg aaaccagttc tgaacaattt 4260
agaaggaaaa ccttgaaaac atttaacaaa aaattatatt ttaatgttta tgaataagag 4320

gaggcttttg aaaaaatgtt gatctataaa tacttacttt aggcctgagg tgtctaatga 4380
gtgaactgag caatgggaac tcaaggctga agcctcctgc atcagaggag gtagaaccag 4440
gagcctcttg agatttgagg tgttttagca ttggaaagcc actctttggg tagctggccc 4500
cagaaactac ttctgacctt gtcatttgga atggaggtta gtggtctgcc agatgccaaa 4560
gctgcatgag accagctctt ggtttatcaa tttgaacact cagtaaccta gaaggcccag 4620
cacaaagtgt ctgctctctt cttaactgag cctgccccag cactactgca caaattaggg 4680
agggtctact tcctacagag catccctccc tgggcccctt cccatctttt gtactctacc 4740
tacctgacct tcaggatctt ggcacatacg aaatggctgt gtagcaagca ctttggcatg 4800
ccctcctaaa cttaccccag agcctctccc tgcctcctta agccagtctg cctgtcttct 4860
ggggaggtgt tagagcccat agaatggaga ggagaaagaa aagaggaaga ggcaggcagg 4920
tagtaaaaag gctctgggag gaaagacagc ctcctaggct ttgcacaagc aggactcagc 4980
cccttgtggg aactaagtgc catcttggag tttaagaaca tttggacaag ttgcaaatga 5040
cctttgctcc ttgctcctct caccttttat ggggccctgc ttagcactga aagcaaatgc 5100
gctgaaaagg caaagaggtt tggctcctgc ccactgatag tcctttccct gcagtgtttg 5160
tgtgtcaagt ggcaaagctg ttcttcctgg tgactctgat tagatccagt aacttaagag 5220
atttgtatgc ataggtctgc tttgactctt ctattctggg cttttgattt gtttttcagt 5280
tttgctttta gttttcctat ttttatttta tgcaccaact agacacacaa agcagttgaa 5340
tttatatata tatatatata tatatatctg tatatttcac aattataaac tcattttgct 5400
tgtgacgcca cacacacaca aaaagaaaaa ccttttaaaa ttatacctgt tgcttaatta 5460
caatatttct gataaccata gagtaggaca agggaaaaaa tttaaaaaaa aaaaaaaaaa 5520
aagaaaaaac acatctgtct gctggtcact tcttcaatcc aagcagatct gtgatctttc 5580
ctcgcgtctt tcaaagactt ccctgtgcta agtgaaggaa gctccaggct gcacccaggt 5640
tttgtgcttt gtttctcctc tgttgtgaaa ggggccccaa gattctgggt acaggacagt 5700
tcatttcagc atggggtcag gagacaagag cactcccttt acatgctgac gtacagaact 5760
tagtgggaat agcctagtcc ccacctctag ggatggggag ctagcatgca tgggggtgac 5820
ccaactccct ccacctttcc ctggccagga agagcctgtg tacagtaagt ctgacaagct 5880
ttccccagtt agcagggctc agagcattta aaaaccctcc aaactttgct gagtctaggg 5940
actagagaga agatagaaga tttggtctat ctccaaggtg tgtaagctgt accaggtaga 6000
atgccaggga ccccagaacc acatccaaca gcccaatggg tctcctccag aaagtagtga 6060

agactccaga aacatccctt tctcttctcc ctgctcccat gagtaactgc atttgctttt 6120
gtaatcctta atgagcatta tctgctaaaa aaaaaaaatt agctgtaaca gttcttttg 6180
caaaaggatc attcttaaat aattaaaaac accccccccc caaaaaaaag tccagaacct 6240
tgttcttcca aagcagagag cattataatc agggccaaaa tctgtcccac acctctaccc 6300
catctcctca tgattgctgc ttctaaggcc agaatacagc aaagatattt gtaggccctt 6360
tgggtgactg ggctaccctt ggagctcttg gaagatgggc tggggaagcc tctgagaccc 6420
tatcctaggg ccttgctcta gggagtaatc agtattagta gagtgtcaca acattattcc 6480
ccagccggca tgagatgggg gcagaagaag ccaaagggtt gtctccactg ctacttactt 6540
ggccactgac aggtaggtga ccatgtatgt ccatatgcat gttttatggc tgatgtgaga 6600
tcagcaccca agttagcttc acctggtgac ctctaaccct gcctggatgg agcaggccac 6660
ctggttcaat gtttctgggc agctggacaa tggagtgcaa aaggcttaca gaacttgaag 6720
ccttttcctt actttgctag cacggcctcc ttttccattt gatttgtcac tgcttcagtc 6780
aataacagcc gctccagagt cagtagttga tgaatatatg accaaatatc accaggactg 6840
ttactcaacg tgtgccgagc cctttccttg tgctgggctc cctgtgtacc tggacactgt 6900
aatgtgtgct gtgtttgctc tccttcctct tccttccttg ccctttcctt gtctttctgg 6960
ggtttttctg ttgggtttgg tttggtttta tttttccttt tgtgttccaa acatgaggtt 7020
ttctctactg gtcctcttta actgtggtgt tgaggcttct atttgtgtaa tttttggtgg 7080
gtgaaaggaa ctttgctaag taaatctctt ctgtgtttga aatgaagtct gtattgtaac 7140
tatgtttaaa gtaattgttc cagagacaaa tgcttctagg tacattttca ttacaaacaa 7200
agcatttgaa gggagggaag tggtgaataa gacaagaggg gcaatctgaa ttgatccctg 7260
cccagatcag ccagaagcta ccaaaagtta agcactggtt ttccattcca agtcaagaga 7320
ctgaagctga tgtttgcca ttttcaaagt caaagcaaaa ccagcttttc cacccaatgg 7380
attctttgct tctccttccc agattattac tactgctgta ataatctagg agtgccagga 7440
gggaaaggag tattaacaca gagctgtgct cactgagtat ggaaaggctt ggtctgagtt 7500
ttcaggagga tgacccactg tggacatggg gagaagacag aagataaatt agccgctccc 7560
tgcctaagat acctcttaat agataagtca aggccatgga cattattgtc tacaaggcat 7620
gtttcaaaga catgaccagt caggacactt ctgtcatact ccatgttgcc ccctagtaca 7680
cagtactaat ctgatatctc tgttcccgcc atgcctgggg gataaaatga tagcagagac 7740
tcctttcctt caatgtgatc taattcccaa caaaatctgg gcctgagata ccacctgttt 7800

ctatggcaaa catcctcagt aaagtgttat tctcattgca gattgttcca gcctaatgta 7860

agaggaacag agcagtgttc ccttggagcc tcatgtggac agttctacct gtagtgacca 7920

gttggctata gtagttatta gctggaacaa ccagacaggg tacatgcccc ctccaaaatc 7980

catgttgtac tcccctctgc cagccagggg gggtgagatc tgtagaatag tgcagccagt 8040

gacaagccac cttgtgtttg tcaccagctc aaaaactcat ctaaggttgg gagcaggcag 8100

acaaggcaga gagaaagatc caggacagac ctagctgggc tggaggggtc ttgaaaagcc 8160

ctctgtcgta ttcaccttca gtttttgtgc tttgggacaa ttactttaga aaataagtag 8220

gtcgttttaa aaacaaaata ttgattgctt ttttgtagtg ttcaaaacaa aaggttcttt 8280

gtgtatagcc aaatgactga aagcactgat atatttaaaa acaaaaggca atttattaag 8340

gaaatttgta ccatttcagt aaacctgtct gaatgtacct gtatacgttt caaaaacaca 8400

ccccactgaa cccctgtaac ctatttatta tataaagagt ttgccttata aatttacata 8460

aaaa 8464

<210> 3
<211> 803
<212> DNA
<213> Hamster

<400> 3
ttgctgcaga tactactgac cagacaagct gttgaccagg cacccccca atactccccc 60

aatgtgctca ttagagatag cagttgagag gacactccca tttttggtgc cctgtccata 120

gcttccctga ctcttccacc accccaactc ccaatctgag ggaccgggag gtgcgaggca 180

ggaaaaatat tggattcttt agagaagact agaggtgacc agtgactgtg gcccagtaat 240

tagaactgtg gtggcacaag tctggcccca catccaccca atccaaaact gataaggata 300

ttttgaaaaa caggaaagca gtacctgtct gatccagctc tggtataggt aggagtgagt 360

cctgaactgc tggattacag actggcttga gccacagaag atgatggacc agagtaaagt 420

atcatcacct gctcacaagg catgcttcac tagagaataa ttctaaagag gtgccatgga 480

ggcagcagga caaggcacaa gcagtctggg tgggggtcaa gccagaccta gtgccacaga 540

acaagagagc aatctgtgac tagtagttag ggactttgtg gatgggacaa ggggcatggg 600

ggaagaaatg aaaatattct tccaattact ttccagttct cctttaggga cagcttagaa 660

ttatttgcac tattgagtct tcatgttccc acttaaaaac aaacagatgc tctgaaagca 720

aactggcttg aaatggtgac actttgtccc acaagccacc aaatgtggca gtgtttagaa 780

ctacctggat ctgtatatac ctg 803

<210> 4
<211> 790
<212> DNA
<213> Kangaroo

<400> 4
ttgctgcata tactactgac cagacaagct gtttatcagg ctttttaggg tacaccagca 60

cctgccctcc attcatccct gttgggagag ggatggtgta ctggttgtca ctagagacct 120

aacagagtag ggttagtggg agcttacatt ttcagtgcca ttaacattct agtccaaggt 180

cttaaattat tatgttgagg ggtttttttt cccctgaggg ggccgggggg tgggggggag 240

gttgattaga ttccttagga aagagggttg agacagacag cagagcactg agcagttggc 300

actaaaggag accttgacta ggggccaggt ggcatcatct aatcccaagg ggctccaagt 360

gagtattagg gtgggggaag acattataga aggaatagaa acaggatagc tcagcctaaa 420

gaagagcggt taaaacccta cccaccagga gttgacttga aagaggcccc tatggaggaa 480

tccccaacca ccaaaagcaa tcttgagctg cagctgcttc atttagtgga ccttgtgtat 540

atctgggtgt gtatgcacat agatagacag tgagaaagaa aactgttctt ccagttcttt 600

tccagtgcta ctagcttagg gacaggttag aactgtctgc acaattgtgt gatcattccc 660

attcccactt caaaacaaac tgactgagat gttcaacaga aaactggctt caatgggtaa 720

catgcccttg ccacttactt aagacactgg tgtgatgggg ttttgaactc cctatatttg 780

taggtatctg 790

<210> 5
<211> 841
<212> DNA
<213> Macaca

<400> 5
ttgctgcaga tactactgac cagacaagct gttgaccagg cacctcccct cccgcccaaa 60

cctttccccc atgtggtcgt tagagacaga cgagttgaga ggacactccc gttttcggtg 120

ccatcagtgc cccgtctacc actcccccag ctcccccact ctcccccact cccaaccacg 180

ttgggacagg gaggtgtgag gcaggagaga cagttggatt ctttagagat ggatgtgacc 240

agtggctatg gcccgtgcga tcccacccgt ggcggctcaa atctggcccc accccagccc 300

caatccaaaa ctggcaagga cgcttcacag gacaggaaag tggcacctgt ctgttccggc 360

atggctagga gggagttgtc ccttgaacta ctgggtgtag actggcctaa atcacaggag 420

aggatggccc agggtgaggt ggcatggtcc attctcaagg gacgtcctcc agttggtggc 480

actagagagg ccatggaggc agtaggacaa ggcacaggca ggctggccca gggtcaggcc 540

gggccgaaca cagcggggtg agagggattc ctcgtctcag agcagtctgt gaccggtagt 600

tagggactta gtggacaggg aaggggcaaa gggggaggag aagaaaatgt tcttccagtt 660

actttccaat tctactcctt tagggacagc ttagaattat ttgcactatt gagtcttcat 720

gttcccactt caaaacaaac agatgctctg agagcaaact ggcttgaatt ggtgacgttt 780

agtccctcag gccaccagat gtgatggtgt tgagaactac ctggatatgt atatatacct 840

g 841

<210> 6
<211> 846
<212> DNA
<213> Orangutan

<400> 6
ttgctgcaga tactactgac cagacaagct gttgaccagg cacctcccct cccgcccaaa 60

cctttccccc atgtggtcgt tagagacaga gcagttgaga ggacactccc gttttcggtg 120

ccatcagtgc cccgtctgca gctcccccag ctccccccac ctccccact cccaaccacg 180

ttgggacagg gaggtgtgag gcaggagaga cagttggatt ctttcgagaa gatggatatg 240

accagtggcc atggcctgtg cgatcccacc cgtggcggct caagtctggc cccacaccag 300

ccccaatcca aaactggcaa ggacgcttca caggacagga aagtggcacc tgtctgctcc 360

agctctggca tggctaggag ggagtcgtcc cttgaactac tgggtgtaga ctggcctgaa 420

ccacaggaga ggatggccca gggtgaggtg gcatggtcca ttctcaaggg acgtcctcca 480

acgggtggcg ctagaaaggc catggaggca gtaggacaag gcgcaggcag gctggcccgg 540

ggtcaggccg ggcagggcac agcggggtga gagggattcc taatcactca gagcagtgtg 600

tgactggtag ttagggactc agtggacagg ggaggggcga ggggcagga gaagaaaatg 660

ttcttccagt tactttccaa ttctccttta gggacagctt agaattattt gcactattga 720

gtcttcatgt tcccacttca aaacaaacga tgctctgaga gcaaactggc ttgaattggt 780

gacatttagt ccctcaagcc accagatgtg agtgttgaga actacctgga tttgtatata 840

tacctg 846

<210> 7
<211> 813
<212> DNA
<213> Rat

<400> 7
ttgctgcaga tactactgac cagacaagct gttgaccagg cactccccac aacaacaacc 60

ccctccctcc tcaccccacc cctatcccct gtgtgctcat tagagagggc aattgagagg 120

acactcccat ttttggtgcc actgatgccc tgtccatagc ttccctgact tttacaccac 180

cccaactccc aatctgaggg actgggaggt gtgacgcagg agaaactata taggactctt 240

gggagaagac tatagagttg gcaagtgatt gcgccccagt aattccaact gtggtagcac 300

aagtctggct ccacaccaac ccaatccaaa actgacaagg acattttgca aaaaatgaaa 360

gtggcatttg tctgatccag ctctggcatg gctagagatg agtcttaaac tgttggctta 420

taaactggcc tgagcaacag aagaggatgg cccagagtaa agtgtcatca tctgttcaca 480

aggcatgctc ccctagaagt tcatgctaaa gaagtgccat ggaggcagca ggacaaagta 540

caggctaggt ggagtcaagc caggcctagt gccacagagc aagagagcag tctctgacta 600

gtagttaagg gggaagaaag aaaaatattc ttccaattgc tttccagttc tcctttaggg 660

acagcttaga attatttgca ctattgagtc ttcatgttcc cacttcaaaa caaatagatg 720

ctctgaaagc aaactggctt gaaatggtga cactgtccca caagccacca gacaatggca 780

gtgttcagaa ctacctgtat atgtatatac ctg 813

<210> 8
<211> 842
<212> DNA
<213> Chimpanzee

<400> 8
ttgctgcaga tactactgac cagacaagct gttgaccagg cacctcccct cccgcccaaa 60

cctttccccc atgtggtcgt tagagacaga gcgacagagc agttgagagg acactcccgt 120

tttcggtgcc atcagtgccc cgtctacagc tcccccagct ccccccacct cccccactcc 180

caaccacgtt gggacaggga ggtgtgaggc aggagagaca gttggattct ttagagaaga 240

tggatatgac cagtggctat ggcctgtgtg atcccacccg tggtggctca agtctggccc 300

cacaccagcc ccaatccaaa actggcaagg acgcttcaca ggacaggaaa gtggcacctg 360

tctgctccag ctctggcatg gctaggaggg gggagtccct tgaactactg ggtgtagact 420

ggcctgaacc acaggagagg atggcccagg gtgaggtggc gtggtccatt ctcaagggac 480

gtcctccaac gggtggcgct agaggccatg gaggcagtag gacaaggcgc aggcaggctg 540

gcccggggtc aggccgggca gagcacagcg gggtgagagg gattcctaat cactcagagc 600

agtctgtgac ttagtggaca ggggagggga caaaggggga ggagaagaaa atgttcttcc 660

agttactttc caattctcct ttagggacag cttagaatta tttgcactat tgagtcttca 720

tgttcccact tcaaaacaaa cagatgctct gagagcaaac tggcttgaat tggtgacatt 780

tagtccctca agccaccaga tgtgacagtg ttgagaacta cctggatttg tatatatacc 840

tg 842

(1) GENERAL INDICATIONS:

(i) APPLICANT:
(A) NAME: Deutsches Krebsforschungszentrum
(B) STREET: Im Neuenheimer Feld 280
(C) TOWN: Heidelberg
(E) COUNTRY: Germany
(F) POSTAl CODE: 69120

(ii) TITLE OF THE INVENTION: Modularly Constructed RNA Molecules Having Two Sequence Region Types

(iii) NUMBER OF SEQUENCES: 8

(iv) COMPUTER-READABLE VERSION:
(A) DATA CARRIER: floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SORTWARE: PatentIn Release #1.0, version #1.30 (EPO)

(v) DATA OF THE CURRENT APPLICATION: not yet known

(vi) DATA OF THE PRIOR APPLICATION:
APPLICATION NUMBER: DE 198 28 624.4
FILING DATE: June 26, 1998

(2) INDICATIONS AS TO ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8422 base pairs
(B) KIND: nucleotide
(C) STRAND FORM: not known
(D) TOPOLOGY: not known

(ii) KIND OF MOLECULE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CTTAGAGTTT CGTGGCTTCA GGGTGGGAGT AGTTGGAGCA TTGGGGATGT TTTTCTTACC 60

GACAAGCACA GTCAGGTTGA AGACCTAACC AGGGCCAGAA GTAGCTTTGC ACTTTTCTAA 120

ACTAGGCTCC TTCAACAAGG CTTGCTGCAG ATACTACTGA CCAGACAAGC TGTTGACCAG 180

GCACCTCCCC TCCCGCCCAA ACCTTTCCCC CATGTGGTCG TTAGAGACAG AGCGACAGAG 240

CAGTTGAGAG GACACTCCCG TTTTCGGTGC CATCAGTGCC CCGTCTACAG CTCCCCCAGC 300

TCCCCCCACC TCCCCCACTC CCAACCACGT TGGGACAGGG AGGTGTGAGG CAGGAGAGAC 360

AGTTGGATTC TTTAGAGAAG ATGGATATGA CCAGTGGCTA TGGCCTGTGC GATCCCACCC 420

GTGGTGGCTC AAGTCTGGCC CCACACCAGC CCCAATCCAA AACTGGCAAG GACGCTTCAC 480

AGGACAGGAA AGTGGCACCT GTCTGCTCCA GCTCTGGCAT GGCTAGGAGG GGGGAGTCCC 540

TTGAACTACT GGGTGTAGAC TGGCCTGAAC CACAGGAGAG GATGGCCCAG GGTGAGGTGG 600

CATGGTCCAT TCTCAAGGGA CGTCCTCCAA CGGGTGGCGC TAGAGGCCAT GGAGGCAGTA 660

GGACAAGGTG CAGGCAGGCT GGCCTGGGGT CAGGCCGGGC AGAGCACAGC GGGGTGAGAG 720

GGATTCCTAA TCACTCAGAG CAGTCTGTGA CTTAGTGGAC AGGGGAGGGG GCAAAGGGGG 780

AGGAGAAGAA AATGTTCTTC CAGTTACTTT CCAATTCTCC TTTAGGGACA GCTTAGAATT 840

ATTTGCACTA TTGAGTCTTC ATGTTCCCAC TTCAAAACAA ACAGATGCTC TGAGAGCAAA 900

CTGGCTTGAA TTGGTGACAT TTAGTCCCTC AAGCCACCAG ATGTGACAGT GTTGAGAACT 960

ACCTGGATTT GTATATATAC CTGCGCTTGT TTTAAAGTGG GCTCAGCACA TAGGGTTCCC 1020

ACGAAGCTCC GAAACTCTAA GTGTTTGCTG CAATTTTATA AGGACTTCCT GATTGGTTTC 1080

TCTTCTCCCC TTCCATTTCT GCCTTTTGTT CATTTCATCC TTTCACTTCT TTCCCTTCCT 1140

CCGTCCTCCT CCTTCCTAGT TCATCCCTTC TCTTCCAGGC AGCCGCGGTG CCCAACCACA 1200

CTTGTCGGCT CCAGTCCCCA GAACTCTGCC TGCCCTTTGT CCTCCTGCTG CCAGTACCAG 1260

CCCCACCCTG TTTTGAGCCC TGAGGAGGCC TTGGGCTCTG CTGAGTCCAA CCTGGCCTGT 1320

CTGTGAAGAG CAAGAGAGCA GCAAGGTCTT GCTCTCCTAG GTAGCCCCCT CTTCCCTGGT 1380

AAGAAAAAGC AAAAGGCATT TCCCACCCTG AACAACGAGC CTTTTCACCC TTCTACTCTA 1440

GAGAAGTGGA CTGGAGGAGC TGGGCCCGAT TTGGTAGTTG AGGAAAGCAC AGAGGCCTCC 1500

TGTGGCCTGC CAGTCATCGA GTGGCCCAAC AGGGGCTCCA TGCCAGCCGA CCTTGACCTC 1560

ACTCAGAAGT CCAGAGTCTA GCGTAGTGCA GCAGGGCAGT AGCGGTACCA ATGCAGAACT 1620

CCCAAGACCC GAGCTGGGAC CAGTACCTGG GTCCCCAGCC CTTCCTCTGC TCCCCCTTTT 1680

CCCTCGGAGT TCTTCTTGAA TGGCAATGTT TTGCTTTTGC TCGATGCAGA CAGGGGGCCA 1740

GAACACCACA CATTTCACTG TCTGTCTGGT CCATAGCTGT GGTGTAGGGG CTTAGAGGCA 1800

TGGGCTTGCT GTGGGTTTTT AATTGATCAG TTTTCATGTG GGATCCCATC TTTTTAACCT 1860

CTGTTCAGGA AGTCCTTATC TAGCTGCATA TCTTCATCAT ATTGGTATAT CCTTTTCTGT 1920

GTTTACAGAG ATGTCTCTTA TATCTAAATC TGTCCAACTG AGAAGTACCT TATCAAAGTA 1980

GCAAATGAGA CAGCAGTCTT ATGCTTCCAG AAACACCCAC AGGCATGTCC CATGTGAGCT 2040

GCTGCCATGA ACTGTCAAGT GTGTGTTGTC TTGTGTATTT CAGTTATTGT CCCTGGCTTC 2100

CTTACTATGG TGTAATCATG AAGGAGTGAA ACATCATAGA AACTGTCTAG CACTTCCTTG 2160

CCAGTCTTTA GTGATCAGGA ACCATAGTTG ACAGTTCCAA TCAGTAGCTT AAGAAAAAAC 2220

CGTGTTTGTC TCTTCTGGAA TGGTTAGAAG TGAGGGAGTT TGCCCCGTTC TGTTTGTAGA 2280

GTCTCATAGT TGGACTTTCT AGCATATATG TGTCCATTTC CTTATGCTGT AAAAGCAAGT 2340

CCTGCAACCA AACTCCCATC AGCCCAATCC CTGATCCCTG ATCCCTTCCA CCTGCTCTGC 2400

TGATGACCCC CCCAGCTTCA CTTCTGACTC TTCCCCAGGA AGGGAAGGGG GGTCAGAAGA 2460

GAGGGTGAGT CCTCCAGAAC TCTTCCTCCA AGGACAGAAG GCTCCTGCCC CCATAGTGGC 2520

CTCGAACTCC TGGCACTACC AAAGGACACT TATCCACGAG AGCGCAGCAT CCGACCAGGT 2580

TGTCACTGAG AAGATGTTTA TTTTGGTCAG TTGGGTTTTT ATGTATTATA CTTAGTCAAA 2640

TGTAATGTGG CTTCTGGAAT CATTGTCCAG AGCTGCTTCC CCGTCACCTG GGCGTCATCT 2700

GGTCCTGGTA AGAGGAGTGC GTGGCCCACC AGGCCCCCCT GTCACCCATG ACAGTTCATT 2760

CAGGGCCGAT GGGGCAGTCG TGGTTGGGAA CACAGCATTT CAAGCGTCAC TTTATTTCAT 2820

TCGGGCCCCA CCTGCAGCTC CCTCAAAGAG GCAGTTGCCC AGCCTCTTTC CCTTCCAGTT 2880

TATTCCAGAG CTGCCAGTGG GGCCTGAGGC TCCTTAGGGT TTTCTCTCTA TTTCCCCCTT 2940

TCTTCCTCAT TCCCTCGTCT TTCCCAAAGG CATCACGAGT CAGTCGCCTT TCAGCAGGCA 3000

GCCTTGGCGG TTTATCGCCC TGGCAGGCAG GGGCCCTGCA GCTCTCATGC TGCCCCTGCC 3060

TTGGGGTCAG GTTGACAGGA GGTTGGAGGG AAAGCCTTAA GCTGCAGGAT TCTCACCAGC 3120

TGTGTCCGGC CCAGTTTTGG GGTCTGACCT CAATTTCAAT TTTGTCTGTA CTTGAACATT 3180

ATGAAGATGG GGGCCTCTTT CAGTGAATTT GTGAACAGCA GAATTGACCG ACAGCTTTCC 3240

AGTACCCATG GGGCTAGGTC ATTAAGGCCA CATCCACAGT CTCCCCCACC CTTGTTCCAG 3300

TTGTTAGTTA CTACCTCCTC TCCTGACAAT ACTGTATGTC GTCGAGCTCC CCCCAGGTCT 3360

ACCCCTCCCG GCCCTGCCTG CTGGTGGGCT TGTCATAGCC AGTGGGATTG CCGGTCTTGA 3420

CAGCTCAGTG AGCTGGAGAT ACTTGGTCAC AGCCAGGCGC TAGCACAGCT CCCTTCTGTT 3480

GATGCTGTAT TCCCATATCA AAAGGCACAG GGGACACCCA GAAACGCCAC ATCCCCCAAT 3540

CCATCAGTGC CAAACTAGCC AACGGCCCCA GCTTCTCAGC TCGCTGGATG GCGGAAGCTG 3600

CTACTCGTGA GCGCCAGTGC GGGTGCAGAC AATCTTCTGT TGGGTGGCAT CATTCCAGGC 3660

CCGAAGCATG AACAGTGCAC CTGGGACAGG GAGCAGCCCC AAATTGTCAC CTGCTTCTCT 3720

GCCCAGCTTT TCATTGCTGT GACAGTGATG GCGAAAGAGG GTAATAACCA GACACAAACT 3780

GCCAAGTTGG GTGGAGAAAG GAGTTTCTTT AGCTGACAGA ATCTCTGAAT TTTAAATCAC 3840

TTAGTAAGCG GCTCAAGCCC AGGAGGGAGC AGAGGGATAC GAGCGGAGTC CCCTGCGCGG 3900

GACCATCTGG AATTGGTTTA GCCCAAGTGG AGCCTGACAG CCAGAACTCT GTGTCCCCCG 3960

TCTAACCACA GCTCCTTTTC CAGAGCATTC CAGTCAGGCT CTCTGGGCTG ACTGGGCCAG 4020

GGGAGGTTAC AGGTACCAGT TCTTTAAGAA GATCTTTGGG CATATACATT TTTAGCCTGT 4080

GTCATTGCCC CAAATGGATT CCTGTTTCAA GTTCACACCT GCAGATTCTA GGACCTGTGT 4140

CCTAGACTTC AGGGAGTCAG CTGTTTCTAG AGTTCCTACC ATGGAGTGGG TCTGGAGGAC 4200

CTGCCCGGTG GGGGGGCAGA GCCCTGCTCC CTCCGGGTCT TCCTACTCTT CTCTCTGCTC 4260

TGACGGGATT TGTTGATTCT CTCCATTTTG GTGTCTTTCT CTTTTAGATA TTGTATCAAT 4320

CTTTAGAAAA GGCATAGTCT ACTTGTTATA AATCGTTAGG ATACTGCCTC CCCCAGGGTC 4380

TAAAATTACA TATTAGAGGG GAAAAGCTGA ACACTGAAGT CAGTTCTCAA CAATTTAGAA 4440

GGAAAACCTA GAAAACATTT GGCAGAAAAT TACATTTCGA TGTTTTTGAA TGAATACAAG 4500

CAAGCTTTTA CAACAGTGCT GATCTAAAAA TACTTAGCAC TTGGCCTGAG ATGCCTGGTG 4560

AGCATTACAG GCAAGGGGAA TCTGGAGGTA GCCGACCTGA GGACATGGCT TCTGAACCTG 4620

TCTTTTGGGA GTGGTATGGA AGGTGGAGCG TTCACCAGTG ACCTGGAAGG CCCAGCACCA 4680

CCCTCCTTCC CACTCTTCTC ATCTTGACAG AGCCTGCCCC AGCGCTGACG TGTCAGGAAA 4740

ACACCCAGGG AACTAGGAAG GCACTTCTGC CTGAGGGGCA GCCTGCCTTG CCCACTCCTG 4800
CTCTGCTCGC CTCGGATCAG CTGAGCCTTC TGAGCTGGCC TCTCACTGCC TCCCCAAGGC 4860
CCCCTGCCTG CCCTGTCAGG AGGCAGAAGG AAGCAGGTGT GAGGGCAGTG CAAGGAGGGA 4920
GCACAACCCC CAGCTCCCGC TCCGGGCTCC GACTTGTGCA CAGGCAGAGC CCAGACCCTG 4980
GAGGAAATCC TACCTTTGAA TTCAAGAACA TTTGGGGAAT TTGGAAATCT CTTTGCCCCC 5040
AAACCCCCAT TCTGTCCTAC CTTTAATCAG GTCCTGCTCA GCAGTGAGAG CAGATGAGGT 5100
GAAAAGGCCA AGAGGTTTGG CTCCTGCCCA CTGATAGCCC CTCTCCCCGC AGTGTTTGTG 5160
TGTCAAGTGG CAAAGCTGTT CTTCCTGGTG ACCCTGATTA TATCCAGTAA CACATAGACT 5220
GTGCGCATAG GCCTGCTTTG TCTCCTCTAT CCTGGGCTTT TGTTTTGCTT TTTAGTTTTG 5280
CTTTTAGTTT TTCTGTCCCT TTTATTTAAC GCACCGACTA GACACACAAA GCAGTTGAAT 5340
TTTTATATAT ATATCTGTAT ATTGCACAAT TATAAACTCA TTTTGCTTGT GGCTCCACAC 5400
ACACAAAAAA AGACCTGTTA AAATTATACC TGTTGCTTAA TTACAATATT TCTGATAACC 5460
ATAGCATAGG ACAAGGGAAA ATAAAAAAAG AAAAAAAAGA AAAAAAAACG ACAAATCTGT 5520
CTGCTGGTCA CTTCTTCTGT CCAAGCAGAT TCGTGGTCTT TTCCTCGCTT CTTTCAAGGG 5580
CTTTCCTGTG CCAGGTGAAG GAGGCTCCAG GCAGCACCCA GGTTTTGCAC TCTTGTTTCT 5640
CCCGTGCTTG TGAAAGAGGT CCCAAGGTTC TGGGTGCAGG AGCGCTCCCT TGACCTGCTG 5700
AAGTCCGGAA CGTAGTCGGC ACAGCCTGGT CGCCTTCCAC CTCTGGGAGC TGGAGTCCAC 5760
TGGGGTGGCC TGACTCCCCC AGTCCCCTTC CCGTGACCTG GTCAGGGTGA GCCCATGTGG 5820
AGTCAGCCTC GCAGGCCTCC CTGCCAGTAG GGTCCGAGTG TGTTTCATCC TTCCCACTCT 5880
GTCGAGCCTG GGGGCTGGAG CGGAGACGGG AGGCCTGGCC TGTCTCGGAA CCTGTGAGCT 5940
GCACCAGGTA GAACGCCAGG GACCCCAGAA TCATGTGCGT CAGTCCAAGG GGTCCCCTCC 6000
AGGAGTAGTG AAGACTCCAG AAATGTCCCT TTCTTCTCCC CCATCCTACG AGTAATTGCA 6060
TTTGCTTTTG TAATTCTTAA TGAGCAATAT CTGCTAGAGA GTTTAGCTGT AACAGTTCTT 6120
TTTGATCATC TTTTTTTAAT AATTAGAAAC ACCAAAAAAA TCCAGAAACT TGTTCTTCCA 6180
AAGCAGAGAG CATTATAATC ACCAGGGCCA AAAGCTTCCC TCCCTGCTGT CATTGCTTCT 6240
TCTGAGGCCT GAATCCAAAA GAAAAACAGC CATAGGCCCT TTCAGTGGCC GGGCTACCCG 6300
TGAGCCCTTC GGAGGACCAG GGCTGGGGCA GCCTCTGGGC CCACATCCGG GGCCAGCTCC 6360
GGCGTGTGTT CAGTGTTAGC AGTGGGTCAT GATGCTCTTT CCCACCCAGC CTGGGATAGG 6420
GGCAGAGGAG GCGAGGAGGC CGTTGCCGCT GATGTTTGGC CGTGAACAGG TGGGTGTCTG 6480
CGTGCGTCCA CGTGCGTGTT TTCTGACTGA CATGAAATCG ACGCCCGAGT TAGCCTCACC 6540
CGGTGACCTC TAGCCCTGCC CGGATGGAGC GGGGCCCACC CGGTTCAGTG TTTCTGGGGA 6600
GCTGGACAGT GGAGTGCAAA AGGCTTGCAG AACTTGAAGC CTGCTCCTTC CCTTGCTACC 6660
ACGGCCTCCT TTCCGTTTGA TTTGTCACTG CTTCAATCAA TAACAGCCGC TCCAGAGTCA 6720
GTAGTCAATG AATATATGAC CAAATATCAC CAGGACTGTT ACTCAATGTG TGCCGAGCCC 6780

TTGCCCATGC TGGGCTCCCG TGTATCTGGA CACTGTAACG TGTGCTGTGT TTGCTCCCCT 6840

TCCCCTTCCT TCTTTGCCCT TTACTTGTCT TTCTGGGGTT TTTCTGTTTG GGTTTGGTTT 6900

GGTTTTTATT TCTCCTTTTG TGTTCCAAAC ATGAGGTTCT CTCTACTGGT CCTCTTAACT 6960

GTGGTGTTGA GGCTTATATT TGTGTAATTT TTGGTGGGTG AAAGGAATTT TGCTAAGTAA 7020

ATCTCTTCTG TGTTTGAACT GAAGTCTGTA TTGTAACTAT GTTTAAAGTA ATTGTTCCAG 7080

AGACAAATAT TTCTAGACAC TTTTTCTTTA CAAACAAAAG CATTCGGAGG GAGGGGGATG 7140

GTGACTGAGA TGAGAGGGGA GAGCTGAACA GATGACCCCT GCCCAGATCA GCCAGAAGCC 7200

ACCCAAAGCA GTGGAGCCCA GGAGTCCCAC TCCAAGCCAG CAAGCCGAAT AGCTGATGTG 7260

TTGCCACTTT CCAAGTCACT GCAAAACCAG GTTTTGTTCC GCCCAGTGGA TTCTTGTTTT 7320

GCTTCCCCTC CCCCCGAGAT TATTACCACC ATCCCGTGCT TTTAAGGAAA GGCAAGATTG 7380

ATGTTTCCTT GAGGGGAGCC AGGAGGGGAT GTGTGTGTGC AGAGCTGAAG AGCTGGGGAG 7440

AATGGGGCTG GGCCCACCCA AGCAGGAGGC TGGGACGCTC TGCTGTGGGC ACAGGTCAGG 7500

CTAATGTTGG CAGATGCAGC TCTTCCTGGA CAGGCCAGGT GGTGGGCATT CTCTCTCCAA 7560

GGTGTGCCCC GTGGGCATTA CTGTTTAAGA CACTTCCGTC ACATCCCACC CCATCCTCCA 7620

GGGCTCAACA CTGTGACATC TCTATTCCCC ACCCTCCCCT TCCCAGGGCA ATAAAATGAC 7680

CATGGAGGGG GCTTGCACTC TCTTGGCTGT CACCCGATCG CCAGCAAAAC TTAGATGTGA 7740

GAAAACCCCT TCCCATTCCA TGGCGAAAAC ATCTCCTTAG AAAAGCCATT ACCCTCATTA 7800

GGCATGGTTT TGGGCTCCCA AAACACCTGA CAGCCCCTCC CTCCTCTGAG AGGCGGAGAG 7860

TGCTGACTGT AGTGACCATT GCATGCCGGG TGCAGCATCT GGAAGAGCTA GGCAGGGTGT 7920

CTGCCCCCTC CTGAGTTGAA GTCATGCTCC CCTGTGCCAG CCCAGAGGCC GAGAGCTATG 7980

GACAGCATTG CCAGTAACAC AGGCCACCCT GTGCAGAAGG GAGCTGGCTC CAGCCTGGAA 8040

ACCTGTCTGA GGTTGGGAGA GGTGCACTTG GGGCACAGGG AGAGGCCGGG ACACACTTAG 8100

CTGGAGATGT CTCTAAAAGC CCTGTATCGT ATTCACCTTC AGTTTTTGTG TTTTGGGACA 8160

ATTACTTTAG AAAATAAGTA GGTCGTTTTA AAAACAAAAA TTATTGATTG CTTTTTTGTA 8220

GTGTTCAGAA AAAAGGTTCT TTGTGTATAG CCAAATGACT GAAAGCACTG ATATATTTAA 8280

AAACAAAAGG CAATTTATTA AGGAAATTTG TACCATTTCA GTAAACCTGT CTGAATGTAC 8340

CTGTATACGT TTCAAAAACA CCCCCCCCCC ACTGAATCCC TGTAACCTAT TTATTATATA 8400

AAGAGTTTGC CTTATAAATT TA 8422

(2) INDICATIONS AS TO ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8464 amino acids
- (B) KIND: nucleotide
- (C) STRAND FORM: not known
- (D) TOPOLOGY: not known

(ii) KIND OF MOLECULE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CTTAGAGTTT CGTGGCTTCG GGGTGGGAGT AGTTGGAGCA TTGGGATGTT TTTCTTACCG 60

ACAAGCACAG TCAGGTTGAA GACCTAACCA GGGCCAGAAG TAGCTTTGCA CTTTTCTAAA 120

CTAGGCTCCT TCAACAAGGC TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG 180

CACTCCCCCC AACAATATCC TCCCTCTTCC CCCCCCCCAC CCCCGCCCCG TGTGCTCGTT 240

AGGGCAATTG AAAGGACACT CCCATTTTTG GTGCCATTGA TGCCCTGTCC ATAATAGCTT 300

CCCTGACTTT TACACCACCC CAACTCCCAA TCTGAAGGAC TGGGAGGTGT GATGCAGGAG 360

AAACTATGGG ACTCTTGGGA GAAGACTATG GAGTTGGCCA GTGATTAAGG CCCACTAATT 420

CCAACTGTGG TAGCACAGAT CTGGCTCCAC ATCAACCCAA TCCAAAACTG ACAAGGATAT 480

TTTGCAAAAA AAGAAAGTGG CACCTGTCTG ATCCAGCTCT GACATGGCTA GAGGTGAGTC 540

CTAAACTGAT GGCTTATAAA CTAGCCTGAG CCACAGAAGA GTATGGCCCA GAGTGAAGTG 600

TCATCATCTG TTCACAAGGC ATGCTCCCCT AGAAGATAAT GCTAAAGAGG TGCCATGGAG 660

GCAGCAGGAC AAAGTACAGG CAGGCTAGGT GGAGTCAAGC CAGGCCTAGT GCCACAGAAC 720

AAGAGAGCAG TCTGACTAGT AATTAAGAGG GAAGAAAGGA AAATATTCTT CCAATTACTT 780

TCCAGTTCTC CTTTAGGGAC AGCTTAGAAT TATTTGCACT ATTGAGTCTT CATGTTCCCA 840

CTTCAAAACA AACAGATGCT CTGAAAGCAA ACTGGCTTGA AATGGTGACA CTGTCCCACA 900

AGCCACCAGA CATGGCAGTG TTCAGAACTA CCTGTATCTG TATATACCTG CGCTTGTTTT 960

AAAGTGGGCT CAGCACATAG GATTCCCAAG AAGCTCCGAA ACTCTAAGTG TTTGCTGCAA 1020

TTTTATAAGG ACTTCCTGAT TGCTTTCTCT CTCGTCCTTC CATTTCTTCC TTCCTTCCAT 1080

TTCATGCTTT CATTTCTTCC CCTAGCTTCT AGTTGTTTCT TCTGTTCCAG GCAGCTGCAG 1140

TGCTGAACCA CATGGTTACC TAACAGCAGT CAGCTGCAGC CCTAGGATTC TTCCTGCCCT 1200

TTAACTTCCC ATTGCCAGTG CCAGGTATCA TATTTAACCT TGAGCAAGAG CTGGGCTCTT 1260

TTGAGCCCTC CCTAACCTCT GTGAAGAAGA ACAAGAAGGT AGGAAGCTCT TGCTCTTGCT 1320

AAGAAAAATG TCAAAAGGCT TTCAGACCTT AAACAATGAG CCTTTTCACC TTTTACTCTA 1380

GAAAAGTGGA CTAGAAAATC TGGGTCACAT TGGGTAGCTG AAGGAGATAC AGAGGCCCCT 1440

ATGGCCTGCC AGAGTCGTTG CATGGCCCAA CAGGGGCTCC ATGCCCACTA CCCTTGACCC 1500

TACTCAGAAA TCTAATGTCA TACTTAGTGT GGGCAGGGGA CCTGTCAGGA CAGATGCAGA 1560

CCTAAGCAGG GAGTGACACC AGGGCCCTTG GCCCTTCTTC TGACAAACAT ACACATCCCA 1620

AGTCTTTTTC TAGTGGAATT CTTAACCTCT TGCTCACTGG GGACTGGGAA GCATCAGCAC 1680

ATCCCATATT TCAAACTCTG CTCCATAAGT ACAGTGGTGA ATTTTATAGA CTTGACTTTG 1740

CTGTGGGGTT TTAATTGGTC AGTTTTAATT TGGGATCCCA AAGTTTTAAC CTCCATTCAG 1800

GAAGTCCTTA TCTAGCTGCA TATCTTCATC ATATTGGTAT ATCCTTTTCT GTGTTTACAG 1860

AGATGTCTCA TATCTATCGA AATCTGTCTG AGAAGTACCT TATCAAAGTA GCAAATGAGA 1920

CAGCAGTCTT ATGCTTCCAG AAACACCCAC AGGCACGTCC CATGTGAGCT GCTGCCATGA 1980

ACTGTCGAGT GTGTATTGTC TTGTGTATTT TCGTTAACGT TCCCCAGCTT CCTTCCTGCG 2040

GTGTAATCAT GGAAGAGTGA AACATCATAG AAATCGTCTA GCACTTCCTG GCCAGTCCTT 2100

AGTGATCAGG AACCGTAGTT GACAGTTCCA ATTGATAGCT TAAGATAAAA CCATGTTTGT 2160

CTCTTATGGA ATGGTTAGAA CTAAGTGAGA GATCTTGCCC CATTCTGTTT GCCGAATCAT 2220

AGTTGGACTT TTAGTGTATT TGTATCCATT TCCTTGTGCT ATAAAAGCAA ACCCTGCAAC 2280

CAGCTTTCTG TCAGGCAGTC CTTTTGCCTG CTCTGCTTTT GATCCTCTTA GTCTTGCTTC 2340

TGGTTCCTCC CTGGAGAGGG AGGAGGGGTC AGAAGAGGAA TTCTGGAGGA TCCAGGATAT 2400

GTCCTTCTGA ACTCCTGCTT CTTCCAGTGA CAAAAGGCCC CTACTGCCCC ACCCCAACCT 2460

GCCCCATGCA CTCCTCTAGG ACACCTTTCC ATACTTTTCA CAACACCTAG CCAGGTTGAC 2520

ACCAAGTTGT TTATTGTGGT CTGCTTGGAA TTTTACCTGT TAGGCTTACT TAGTCCAATC 2580

AAATGGACTC CAAGTTGGGT ATCCCTCATC TTTGGAAGAC AACCTAGGCT GATTAGATAT 2640

TTACTTTTGG GATTGCAGCA CTTTGGGTGC CGTTTTTCTT TTACTTGGGT TTTATCTGCA 2700

GCTCCCTCAC CACCACCACC ACCCCCCACT TACCTGTATG TAGAACTGAT TTCAAAACTG 2760

CAGGTGGTGG TAACTGCAGC TTCTTAGGGT TTTCTTCACT TCTTGCTTCT TTCCCCATTC 2820

CCTCATCCAC AAATAAGGGC ATCACAAGTC AGTCTCCTTT AAGCAGGCAG CTTTGGTGGG 2880

GTTTTTCCCC TGGAAGCCAG GGACCCTGTC AGGCTGCCTC TGCCTTGTGG TCAGGTTGAC 2940

AGGAGGTTGG AGGGAAAAGC CTTAAGTCAT GGGATTCTCA CCAGCTGTGT CTGGCTCAGA 3000

CCTGGAATGT GACCTTTATT TTGTTGTATT TGAACATTGT AAAGTGTGGG TGGTACCTTA 3060

AACTGAATAT GTGAAGAATC CAGAAACTGA CCAACAGCTT TCAGATACCT GGGGCTAGGT 3120

CACTAAGGTC ACATCCAGTC TTCCCTACCC TGTTCTAGTT GTTAGCTACT ACCTCTCCCA 3180

GATAGATTGC TGTATATCCT CCAACTATGA TCATCCTGGC CCAAGCTTGC CTGTTCTTGA 3240

GTCTGTCTTA ACCAGTGGAA CTGCTGCCCT TGGTGTGCAG TGAGTTGAGG ACTCTTGGTC 3300

ACAGCCAGGC TCTAGTAGTA CAGCTCCTTT CTGCTGGTGC TGTATTTCCA TATCAAAAGG 3360

CACAGGGGAG ATCTAGAAAT GCCATCTCCC CCAGTCCATC AGTGCCAAAC AAGCCCATGA 3420

TCCCAGCATG GGTACAGACA ACTCTGTTCA GTGCTATCAC AACAGACTAG AGGCCATGAA 3480

CATTGGACGT GGGAACCAGA GCAACCCGAA TTGCTGCTGC TTTATTCAGC TTTCCGTTGC 3540

TCTGACAATG ATAAAACAAG GCAGTAACTT AAAACAGACT GCCAGGTTTG GCAGAGAAAG 3600

GAAATTCCTT AGCTGACAGC ACCTCTGGAT TTTAAATAGG TTGTAATAAG TGGCTCAAAC 3660

CCATCCAGGA AAAAGCAAAA GGGTTAGAAC TGACCAGATG AGACCAGCCT GATTTCATGC 3720

AGCCCAAATG GAGTCCAGCT GTCTGAACTC TGCAGCACTT CTCTACTACA GTCTCCTAGA 3780

GCATTCCAGC CAGGCTCTTC AGGCTGAGGA GACATCACAG GTGCCAGTTC TTCAAGAAGA 3840

CTTTTGTGCA TCAGTTCATA GCCTATATCT TTGCCCAAGA TTGTAGATTC AGGTTAACAC 3900

TACAGATTCT AGGGCAGATG ACTGAGACTC AGAAAAAAAG CCCCTGTGGA CTGTGGTATA 3960
GCGAAGTACA AAAACTGAAG GGGGCTAGGG CAGATGCCGC ATGCCTCATG CCAGAGCCAA 4020
GCCCTCTGCT CCATCCACAT CCTTTTCTGG CTCCTTCTTC CTGCTCTCTG CTTCAGTGAA 4080
CCAGCCCCAC TCTGAAGAGA TTTGTTGATT CTCTCCATTT TTATGTCTTT CTCTTTTAGG 4140
TACTATATAG AAAAGGCTTA GTCTAATTGT TATAAATTGC TAGAATACTG CCTCCCCCAG 4200
GGTCTAAAAA TATATGCTAA AGGGGAAAAC TTGAACACTG AAACCAGTTC TGAACAATTT 4260
AGAAGGAAAA CCTTGAAAAC ATTTAACAAA AAATTATATT TTAATGTTTA TGAATAAGAG 4320
GAGGCTTTTG AAAAAATGTT GATCTATAAA TACTTACTTT AGGCCTGAGG TGTCTAATGA 4380
GTGAACTGAG CAATGGGAAC TCAAGGCTGA AGCCTCCTGC ATCAGAGGAG GTAGAACCAG 4440
GAGCCTCTTG AGATTTGAGG TGTTTTAGCA TTGGAAAGCC ACTCTTTGGG TAGCTGGCCC 4500
CAGAAACTAC TTCTGACCTT GTCATTTGGA ATGGAGGTTA GTGGTCTGCC AGATGCCAAA 4560
GCTGCATGAG ACCAGCTCTT GGTTTATCAA TTTGAACACT CAGTAACCTA GAAGGCCCAG 4620
CACAAAGTGT CTGCTCTCTT CTTAACTGAG CCTGCCCCAG CACTACTGCA CAAATTAGGG 4680
AGGGTCTACT TCCTACAGAG CATCCCTCCC TGGGCCCCCT CCCATCCTTT GTACTCTACC 4740
TACCTGACCT TCAGGATCTT GGCACATACG AAATGGCTGT GTAGCAAGCA CTTTGGCATG 4800
CCCTCCTAAA CTTACCCCAG AGCCTCTCCC TGCCTCCTTA AGCCAGTCTG CCTGTCTTCT 4860
GGGGAGGTGT TAGAGCCCAT AGAATGGAGA GGAGAAAGAA AAGAGGAAGA GGCAGGCAGG 4920
TAGTAAAAAG GCTCTGGGAG GAAAGACAGC CTCCTAGGCT TTGCACAAGC AGGACTCAGC 4980
CCCTTGTGGG AACTAAGTGC CATCTTGGAG TTTAAGAACA TTTGGACAAG TTGCAAATGA 5040
CCTTTGCTCC TTGCTCCTCT CACCTTTTAT GGGGCCCTGC TTAGCACTGA AAGCAAATGC 5100
GCTGAAAAGG CAAAGAGGTT TGGCTCCTGC CCACTGATAG TCCTTTCCCT GCAGTGTTTG 5160
TGTGTCAAGT GGCAAAGCTG TTCTTCCTGG TGACTCTGAT TAGATCCAGT AACTTAAGAG 5220
ATTTGTATGC ATAGGTCTGC TTTGACTCTT CTATTCTGGG CTTTTGATTT GTTTTTCAGT 5280
TTTGCTTTTA GTTTTCCTAT TTTTATTTTA TGCACCAACT AGACACACAA AGCAGTTGAA 5340
TTTATATATA TATATATATA TATATATCTG TATATTTCAC AATTATAAAC TCATTTTGCT 5400
TGTGACGCCA CACACACACA AAAAGAAAAA CCTTTTAAAA TTATACCTGT TGCTTAATTA 5460
CAATATTTCT GATAACCATA GAGTAGGACA AGGGAAAAAA TTTAAAAAAA AAAAAAAAAA 5520
AAGAAAAAAC ACATCTGTCT GCTGGTCACT TCTTCAATCC AAGCAGATCT GTGATCTTTC 5580
CTCGCGTCTT TCAAAGACTT CCCTGTGCTA AGTGAAGGAA GCTCCAGGCT GCACCCAGGT 5640
TTTGTGCTTT GTTTCTCCTC TGTTGTGAAA GGGGCCCCAA GATTCTGGGT ACAGGACAGT 5700
TCATTTCAGC ATGGGGTCAG GAGACAAGAG CACTCCCTTT ACATGCTGAC GTACAGAACT 5760
TAGTGGGAAT AGCCTAGTCC CCACCTCTAG GGATGGGGAG CTAGCATGCA TGGGGGTGAC 5820
CCAACTCCCT CCACCTTTCC CTGGCCAGGA AGAGCCTGTG TACAGTAAGT CTGACAAGCT 5880
TTCCCCAGTT AGCAGGGCTC AGAGCATTTA AAAACCCTCC AAACTTTGCT GAGTCTAGGG 5940

ACTAGAGAGA AGATAGAAGA TTTGGTCTAT CTCCAAGGTG TGTAAGCTGT ACCAGGTAGA 6000

ATGCCAGGGA CCCCAGAACC ACATCCAACA GCCCAATGGG TCTCCTCCAG AAAGTAGTGA 6060

AGACTCCAGA AACATCCCTT TCTCTTCTCC CTGCTCCCAT GAGTAACTGC ATTTGCTTTT 6120

GTAATCCTTA ATGAGCATTA TCTGCTAAAA AAAAAAAATT AGCTGTAACA GTTCTTTTTG 6180

CAAAAGGATC ATTCTTAAAT AATTAAAAAC ACCCCCCCCC CAAAAAAAAG TCCAGAACCT 6240

TGTTCTTCCA AAGCAGAGAG CATTATAATC AGGGCCAAAA TCTGTCCCAC ACCTCTACCC 6300

CATCTCCTCA TGATTGCTGC TTCTAAGGCC AGAATACAGC AAAGATATTT GTAGGCCCTT 6360

TGGGTGACTG GGCTACCCTT GGAGCTCTTG GAAGATGGGC TGGGGAAGCC TCTGAGACCC 6420

TATCCTAGGG CCTTGCTCTA GGGAGTAATC AGTATTAGTA GAGTGTCACA ACATTATTCC 6480

CCAGCCGGCA TGAGATGGGG GCAGAAGAAG CCAAAGGGTT GTCTCCACTG CTACTTACTT 6540

GGCCACTGAC AGGTAGGTGA CCATGTATGT CCATATGCAT GTTTTATGGC TGATGTGAGA 6600

TCAGCACCCA AGTTAGCTTC ACCTGGTGAC CTCTAACCCT GCCTGGATGG AGCAGGCCAC 6660

CTGGTTCAAT GTTTCTGGGC AGCTGGACAA TGGAGTGCAA AAGGCTTACA GAACTTGAAG 6720

CCTTTTCCTT ACTTTGCTAG CACGGCCTCC TTTTCCATTT GATTTGTCAC TGCTTCAGTC 6780

AATAACAGCC GCTCCAGAGT CAGTAGTTGA TGAATATATG ACCAAATATC ACCAGGACTG 6840

TTACTCAACG TGTGCCGAGC CCTTTCCTTG TGCTGGGCTC CCTGTGTACC TGGACACTGT 6900

AATGTGTGCT GTGTTTGCTC TCCTTCCTCT TCCTTCCTTG CCCTTTCCTT GTCTTTCTGG 6960

GGTTTTTCTG TTGGGTTTGG TTTGGTTTTA TTTTTCCTTT TGTGTTCCAA ACATGAGGTT 7020

TTCTCTACTG GTCCTCTTTA ACTGTGGTGT TGAGGCTTCT ATTTGTGTAA TTTTTGGTGG 7080

GTGAAAGGAA CTTTGCTAAG TAAATCTCTT CTGTGTTTGA AATGAAGTCT GTATTGTAAC 7140

TATGTTTAAA GTAATTGTTC CAGAGACAAA TGCTTCTAGG TACATTTTCA TTACAAACAA 7200

AGCATTTGAA GGGAGGGAAG TGGTGAATAA GACAAGAGGG GCAATCTGAA TTGATCCCTG 7260

CCCAGATCAG CCAGAAGCTA CCAAAAGTTA AGCACTGGTT TTCCATTCCA AGTCAAGAGA 7320

CTGAAGCTGA TGTTTTGCCA TTTTCAAAGT CAAAGCAAAA CCAGCTTTTC CACCCAATGG 7380

ATTCTTTGCT TCTCCTTCCC AGATTATTAC TACTGCTGTA ATAATCTAGG AGTGCCAGGA 7440

GGGAAAGGAG TATTAACACA GAGCTGTGCT CACTGAGTAT GGAAAGGCTT GGTCTGAGTT 7500

TTCAGGAGGA TGACCCACTG TGGACATGGG GAGAAGACAG AAGATAAATT AGCCGCTCCC 7560

TGCCTAAGAT ACCTCTTAAT AGATAAGTCA AGGCCATGGA CATTATTGTC TACAAGGCAT 7620

GTTTCAAAGA CATGACCAGT CAGGACACTT CTGTCATACT CCATGTTGCC CCCTAGTACA 7680

CAGTACTAAT CTGATATCTC TGTTCCCGCC ATGCCTGGGG GATAAAATGA TAGCAGAGAC 7740

TCCTTTCCTT CAATGTGATC TAATTCCCAA CAAAATCTGG GCCTGAGATA CCACCTGTTT 7800

CTATGGCAAA CATCCTCAGT AAAGTGTTAT TCTCATTGCA GATTGTTCCA GCCTAATGTA 7860

AGAGGAACAG AGCAGTGTTC CCTTGGAGCC TCATGTGGAC AGTTCTACCT GTAGTGACCA 7920

GTTGGCTATA GTAGTTATTA GCTGGAACAA CCAGACAGGG TACATGCCCC CTCCAAAATC 7980

CATGTTGTAC TCCCCTCTGC CAGCCAGGGG GGGTGAGATC TGTAGAATAG TGCAGCCAGT 8040

GACAAGCCAC CTTGTGTTTG TCACCAGCTC AAAAACTCAT CTAAGGTTGG GAGCAGGCAG 8100

ACAAGGCAGA GAGAAAGATC CAGGACAGAC CTAGCTGGGC TGGAGGGGTC TTGAAAAGCC 8160

CTCTGTCGTA TTCACCTTCA GTTTTTGTGC TTTGGGACAA TTACTTTAGA AAATAAGTAG 8220

GTCGTTTTAA AAACAAAATA TTGATTGCTT TTTTGTAGTG TTCAAAACAA AAGGTTCTTT 8280

GTGTATAGCC AAATGACTGA AAGCACTGAT ATATTTAAAA ACAAAAGGCA ATTTATTAAG 8340

GAAATTTGTA CCATTTCAGT AAACCTGTCT GAATGTACCT GTATACGTTT CAAAAACACA 8400

CCCCACTGAA CCCCTGTAAC CTATTTATTA TATAAAGAGT TTGCCTTATA AATTTACATA 8460

AAAA 8464

(2) INDICATIONS AS TO ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 803 base pairs
(B) KIND: nucleotide
(C) STRAND FORM: not known
(D) TOPOLOGY: not known

(ii) KIND OF MOLECULE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCCCCCCA ATACTCCCCC 60

AATGTGCTCA TTAGAGATAG CAGTTGAGAG GACACTCCCA TTTTTGGTGC CCTGTCCATA 120

GCTTCCCTGA CTCTTCCACC ACCCCAACTC CCAATCTGAG GGACCGGGAG GTGCGAGGCA 180

GGAAAAATAT TGGATTCTTT AGAGAAGACT AGAGGTGACC AGTGACTGTG GCCCAGTAAT 240

TAGAACTGTG GTGGCACAAG TCTGGCCCCA CATCCACCCA ATCCAAAACT GATAAGGATA 300

TTTTGAAAAA CAGGAAAGCA GTACCTGTCT GATCCAGCTC TGGTATAGGT AGGAGTGAGT 360

CCTGAACTGC TGGATTACAG ACTGGCTTGA GCCACAGAAG ATGATGGACC AGAGTAAAGT 420

ATCATCACCT GCTCACAAGG CATGCTTCAC TAGAGAATAA TTCTAAAGAG GTGCCATGGA 480

GGCAGCAGGA CAAGGCACAA GCAGTCTGGG TGGGGGTCAA GCCAGACCTA GTGCCACAGA 540

ACAAGAGAGC AATCTGTGAC TAGTAGTTAG GGACTTTGTG GATGGGACAA GGGGCATGGG 600

GGAAGAAATG AAAATATTCT TCCAATTACT TTCCAGTTCT CCTTTAGGGA CAGCTTAGAA 660

TTATTTGCAC TATTGAGTCT TCATGTTCCC ACTTAAAAAC AAACAGATGC TCTGAAAGCA 720

AACTGGCTTG AAATGGTGAC ACTTTGTCCC ACAAGCCACC AAATGTGGCA GTGTTTAGAA 780

CTACCTGGAT CTGTATATAC CTG 803

(2) INDICATIONS AS TO ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 790 base pairs
(B) KIND: nucleotide
(C) STRAND FORM: not known
(D) TOPOLOGY: not known

(ii) KIND OF MOLECULE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TTGCTGCATA TACTACTGAC CAGACAAGCT GTTTATCAGG CTTTTTAGGG TACACCAGCA 60
CCTGCCCTCC ATTCATCCCT GTTGGGAGAG GGATGGTGTA CTGGTTGTCA CTAGAGACCT 120
AACAGAGTAG GGTTAGTGGG AGCTTACATT TTCAGTGCCA TTAACATTCT AGTCCAAGGT 180
CTTAAATTAT TATGTTGAGG GGTTTTTTTT CCCCTGAGGG GGCCGGGGGG TGGGGGGAGG 240
GTTGATTAGA TTCCTTAGGA AAGAGGGTTG AGACAGACAG CAGAGCACTG AGCAGTTGGC 300
ACTAAAGGAG ACCTTGACTA GGGGCCAGGT GGCATCATCT AATCCCAAGG GGCTCCAAGT 360
GAGTATTAGG GTGGGGGAAG ACATTATAGA AGGAATAGAA ACAGGATAGC TCAGCCTAAA 420
GAAGAGCGGT TAAAACCCTA CCCACCAGGA GTTGACTTGA AAGAGGCCCC TATGGAGGAA 480
TCCCCAACCA CCAAAAGCAA TCTTGAGCTG CAGCTGCTTC ATTTAGTGGA CCTTGTGTAT 540
ATCTGGGTGT GTATGCACAT AGATAGACAG TGAGAAAGAA AACTGTTCTT CCAGTTCTTT 600
TCCAGTGCTA CTAGCTTAGG GACAGGTTAG AACTGTCTGC ACAATTGTGT GATCATTCCC 660
ATTCCCACTT CAAAACAAAC TGACTGAGAT GTTCAACAGA AAACTGGCTT CAATGGGTAA 720
CATGCCCTTG CCACTTACTT AAGACACTGG TGTGATGGGG TTTTGAACTC CCTATATTTG 780
TAGGTATCTG 790

(2) INDICATIONS AS TO ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 841 base pairs
(B) KIND: nucleotide
(C) STRAND FORM: not known
(D) TOPOLOGY: not known

(ii) KIND OF MOLECULE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCTCCCCT CCCGCCCAAA 60
CCTTTCCCCC ATGTGGTCGT TAGAGACAGA GCAGTTGAGA GGACACTCCC GTTTTCGGTG 120
CCATCAGTGC CCCGTCTACC ACTCCCCCAG CTCCCCCCAC CTCCCCCACT CCCAACCACG 180
TTGGGACAGG GAGGTGTGAG GCAGGAGAGA CAGTTGGATT CTTTAGAGAT GGATGTGACC 240

AGTGGCTATG GCCCGTGCGA TCCCACCCGT GGCGGCTCAA ATCTGGCCCC ACCCCAGCCC 300

CAATCCAAAA CTGGCAAGGA CGCTTCACAG GACAGGAAAG TGGCACCTGT CTGTTCCGGC 360

ATGGCTAGGA GGGAGTTGTC CCTTGAACTA CTGGGTGTAG ACTGGCCTAA ATCACAGGAG 420

AGGATGGCCC AGGGTGAGGT GGCATGGTCC ATTCTCAAGG GACGTCCTCC AGTTGGTGGC 480

ACTAGAGAGG CCATGGAGGC AGTAGGACAA GGCACAGGCA GGCTGGCCCA GGGTCAGGCC 540

GGGCCGAACA CAGCGGGGTG AGAGGGATTC CTCGTCTCAG AGCAGTCTGT GACCGGTAGT 600

TAGGGACTTA GTGGACAGGG AAGGGGCAAA GGGGGAGGAG AAGAAAATGT TCTTCCAGTT 660

ACTTTCCAAT TCTACTCCTT TAGGGACAGC TTAGAATTAT TTGCACTATT GAGTCTTCAT 720

GTTCCCACTT CAAAACAAAC AGATGCTCTG AGAGCAAACT GGCTTGAATT GGTGACGTTT 780

AGTCCCTCAG GCCACCAGAT GTGATGGTGT TGAGAACTAC CTGGATATGT ATATATACCT 840

G 841

(2) INDICATIONS AS TO ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 846 base pairs
(B) KIND: nucleotide
(C) STRAND FORM: not known
(D) TOPOLOGY: not known

(ii) KIND OF MOLECULE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCTCCCCT CCCGCCCAAA 60

CCTTTCCCCC ATGTGGTCGT TAGAGACAGA GCAGTTGAGA GGACACTCCC GTTTTCGGTG 120

CCATCAGTGC CCCGTCTGCA GCTCCCCCAG CTCCCCCCAC CTCCCCCACT CCCAACCACG 180

TTGGGACAGG GAGGTGTGAG GCAGGAGAGA CAGTTGGATT CTTTCGAGAA GATGGATATG 240

ACCAGTGGCC ATGGCCTGTG CGATCCCACC CGTGGCGGCT CAAGTCTGGC CCCACACCAG 300

CCCCAATCCA AAACTGGCAA GGACGCTTCA CAGGACAGGA AAGTGGCACC TGTCTGCTCC 360

AGCTCTGGCA TGGCTAGGAG GGAGTCGTCC CTTGAACTAC TGGGTGTAGA CTGGCCTGAA 420

CCACAGGAGA GGATGGCCCA GGGTGAGGTG GCATGGTCCA TTCTCAAGGG ACGTCCTCCA 480

ACGGGTGGCG CTAGAAAGGC CATGGAGGCA GTAGGACAAG GCGCAGGCAG GCTGGCCCGG 540

GGTCAGGCCG GGCAGGGCAC AGCGGGGTGA GAGGGATTCC TAATCACTCA GAGCAGTGTG 600

TGACTGGTAG TTAGGGACTC AGTGGACAGG GGAGGGGCGA GGGGGCAGGA GAAGAAAATG 660

TTCTTCCAGT TACTTTCCAA TTCTCCTTTA GGGACAGCTT AGAATTATTT GCACTATTGA 720

GTCTTCATGT TCCCACTTCA AAACAAACGA TGCTCTGAGA GCAAACTGGC TTGAATTGGT 780

GACATTTAGT CCCTCAAGCC ACCAGATGTG AGTGTTGAGA ACTACCTGGA TTTGTATATA 840

TACCTG 846

(2) INDICATIONS AS TO ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 813 base pairs
(B) KIND: nucleotide
(C) STRAND FORM: not known
(D) TOPOLOGY: not known

(ii) KIND OF MOLECULE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACTCCCCAC AACAACAACC 60
CCCTCCCTCC TCACCCCACC CCTATCCCCT GTGTGCTCAT TAGAGAGGGC AATTGAGAGG 120
ACACTCCCAT TTTTGGTGCC ACTGATGCCC TGTCCATAGC TTCCCTGACT TTTACACCAC 180
CCCAACTCCC AATCTGAGGG ACTGGGAGGT GTGACGCAGG AGAAACTATA TAGGACTCTT 240
GGGAGAAGAC TATAGAGTTG GCAAGTGATT GCGCCCCAGT AATTCCAACT GTGGTAGCAC 300
AAGTCTGGCT CCACACCAAC CCAATCCAAA ACTGACAAGG ACATTTTGCA AAAAATGAAA 360
GTGGCATTTG TCTGATCCAG CTCTGGCATG GCTAGAGATG AGTCTTAAAC TGTTGGCTTA 420
TAAACTGGCC TGAGCAACAG AAGAGGATGG CCCAGAGTAA AGTGTCATCA TCTGTTCACA 480
AGGCATGCTC CCCTAGAAGT TCATGCTAAA GAAGTGCCAT GGAGGCAGCA GGACAAAGTA 540
CAGGCTAGGT GGAGTCAAGC CAGGCCTAGT GCCACAGAGC AAGAGAGCAG TCTCTGACTA 600
GTAGTTAAGG GGGAAGAAAG AAAAATATTC TTCCAATTGC TTTCCAGTTC TCCTTTAGGG 660
ACAGCTTAGA ATTATTTGCA CTATTGAGTC TTCATGTTCC CACTTCAAAA CAAATAGATG 720
CTCTGAAAGC AAACTGGCTT GAAATGGTGA CACTGTCCCA CAAGCCACCA GACAATGGCA 780
GTGTTCAGAA CTACCTGTAT ATGTATATAC CTG 813

(2) INDICATIONS AS TO ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 842 base pairs
(B) KIND: nucleotide
(C) STRAND FORM: not known
(D) TOPOLOGY: not known

(ii) KIND OF MOLECULE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TTGCTGCAGA TACTACTGAC CAGACAAGCT GTTGACCAGG CACCTCCCCT CCCGCCCAAA 60
CCTTTCCCCC ATGTGGTCGT TAGAGACAGA GCGACAGAGC AGTTGAGAGG ACACTCCCGT 120

TTTCGGTGCC ATCAGTGCCC CGTCTACAGC TCCCCCAGCT CCCCCCACCT CCCCCACTCC 180

CAACCACGTT GGGACAGGGA GGTGTGAGGC AGGAGAGACA GTTGGATTCT TTAGAGAAGA 240

TGGATATGAC CAGTGGCTAT GGCCTGTGTG ATCCCACCCG TGGTGGCTCA AGTCTGGCCC 300

CACACCAGCC CCAATCCAAA ACTGGCAAGG ACGCTTCACA GGACAGGAAA GTGGCACCTG 360

TCTGCTCCAG CTCTGGCATG GCTAGGAGGG GGGAGTCCCT TGAACTACTG GGTGTAGACT 420

GGCCTGAACC ACAGGAGAGG ATGGCCCAGG GTGAGGTGGC GTGGTCCATT CTCAAGGGAC 480

GTCCTCCAAC GGGTGGCGCT AGAGGCCATG GAGGCAGTAG GACAAGGCGC AGGCAGGCTG 540

GCCCGGGGTC AGGCCGGGCA GAGCACAGCG GGGTGAGAGG GATTCCTAAT CACTCAGAGC 600

AGTCTGTGAC TTAGTGGACA GGGGAGGGGG CAAAGGGGGA GGAGAAGAAA ATGTTCTTCC 660

AGTTACTTTC CAATTCTCCT TTAGGGACAG CTTAGAATTA TTTGCACTAT TGAGTCTTCA 720

TGTTCCCACT TCAAAACAAA CAGATGCTCT GAGAGCAAAC TGGCTTGAAT TGGTGACATT 780

TAGTCCCTCA AGCCACCAGA TGTGACAGTG TTGAGAACTA CCTGGATTTG TATATATACC 840

TG 842